The Future of Work:

Machine Learning and Employment

Linda Perkio

Somerville College

Supervisor: Michael Osborne

Department of Engineering

University of Oxford

1 Abstract

This project investigates the susceptibility of jobs to computerisation and particularly which features of a job determine the probability of computerisation. This is achieved by using Gaussian Process Classification. A set of labelled occupations is used to train and test the model and the effect of using different feature sets is explored. Feature selection in the form of greedy selection is used to find the feature set that achieves the best classification and thereby find the features that are most significant when determining if a job can be computerised. It is concluded that the most important feature is Originality and the best feature set for classifying the data in this project consists of Originality and Service Orientation. Furthermore, experiments are performed using linear embedding methods for feature learning. However, these experiments fail to prove that better classification can be achieved using this method.

2 Acknowledgements

I would like to thank my supervisor Michael Osborne for all his advice and guidance throughout this project. His input has been invaluable from start to finish and this finished report would not have been possible without him.

Furthermore, I want to thank my friends and family who have been there to support me throughout my degree and this project. I am grateful for all their support and for the fact that nobody complained when I had to check on one of my experiments during Christmas celebrations or when I was in a bad mood after realising I would have to run several long programs over again.

Finally, I want to thank the Oxford University Ice Hockey Club, which made my time at Oxford unforgettable and provided me with an activity outside of my studies.

3 Contents

4 Introduction

The fact that machines are replacing jobs can be seen every day. Supermarkets are now utilising self-checkout points to replace the cashiers, people get cash out from an ATM instead of a bank office and airlines encourage passengers to check in online rather than at the desk at the airport. This project aims to improve the understanding of how susceptible jobs are to computerisation and in particular exploring what features determine whether a job can be automated.

This analysis builds on the paper “The Future of Employment” (Frey & Michael A. Osborne, 2013), where the susceptibility of jobs to computerisation is investigated. In the paper they use Gaussian Process Classification and information about job features to predict the probability that different jobs can be computerised. This project builds on their work and expands to include more features and explore combinations of features to determine the set of features that best predicts the probability of computerisation. This will help understand the relationship between the job features and the automability of occupations. This is worthwhile because understanding the features that affect if a job can be computerised or not will help understanding the structural changes in the future labour market as current occupations are replaced by machines and new occupations are needed. The skills and abilities identified in this paper as determining if a job can be computerised will be particularly important skills in the future because these are the tasks that are difficult to automate.

In order to determine the importance of different features a set of occupations labelled as possible or impossible to automate are used to train and test a Gaussian Process Classifier using different feature sets as inputs. The classification performance is then evaluated to determine what input data best predicts the probability of computerisation. The features best predicting this probability represent the skills and abilities that cannot be easily automated.

5 Technology and Employment

A changing labour market where technology can replace workers is nothing new. In 1930 Keynes wrote and warned people about the “disease” called “technological unemployment” (Keynes, 1930). Traditionally, routine tasks have been at risk of being replaced by technology but non-routine work has been considered safe from computerisation (Autor, Levy, & Murnane, The Skill Content of Recent Technological Change: An Empirical Exploration, 2003). However, with development in machine learning, mobile robotics and big data computers are capable of doing more and more tasks that have previously seemed impossible. For example, self-driving cars have been developed, a computer beat a the best human team in the quiz show “Jeopardy” and a modern phone has more computing power than any computer in the world had 20 years ago (The Economist, 2014). These are some examples of tasks that are non-routine but have now been successfully computerised. The number of tasks possible to automate is increasing and this means machines will be able to perform more jobs in the future.

5.1 Developments in Technology

The development in technology has had a significant impact on human development. Figure 5.1 shows a Graph of human development over time and as can be seen there is a constant slow rise until the industrial revolution and the development of technology. The graph shows how significant the development of technology has been in the history of human development.

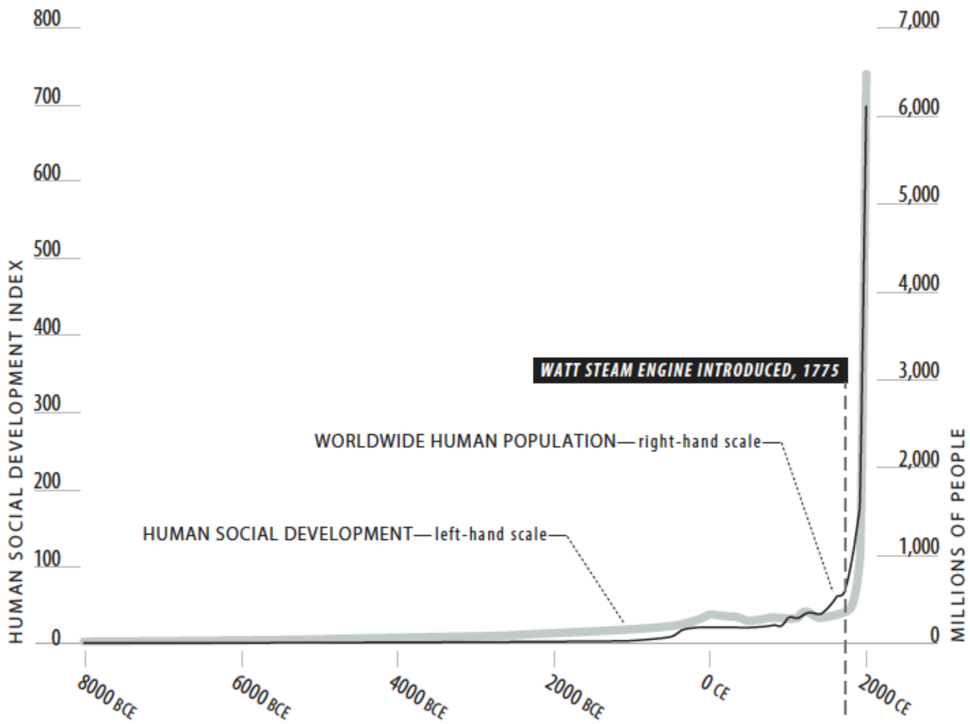

Figure 5.1: Graph showing Human Social Development over time (Brynjolfsson & McAfee, The Second Machine Age, 2014, p. 7)

Brynjolfsson and McAfee argue in “The Second Machine Age” that the development of technology has been the most significant event in human history. They refer to the industrial revolution as the start of the “Machine Age”. Furthermore, they make the argument that we are now entering a second machine age, which will have just as much impact on human development. This highlights the importance of understanding the advantages and limitations of new technology.

5.2 Trends in Employment

Normally it is assumed that economic growth will also mean reduced unemployment. However, after the recession we have seen a so called jobless recovery (Jaimovich & Siu, 2012), meaning that as the economic growth recovered employment has not gone up to the levels before the recession. Figure 5.2 shows the development in GDP per capita and

Labour Force Participation over time in the United States. As seen from the graph even though GDP has showed considerable growth during the last decade the employment rate has fallen. In short, not everyone has been able to benefit from the economic growth.

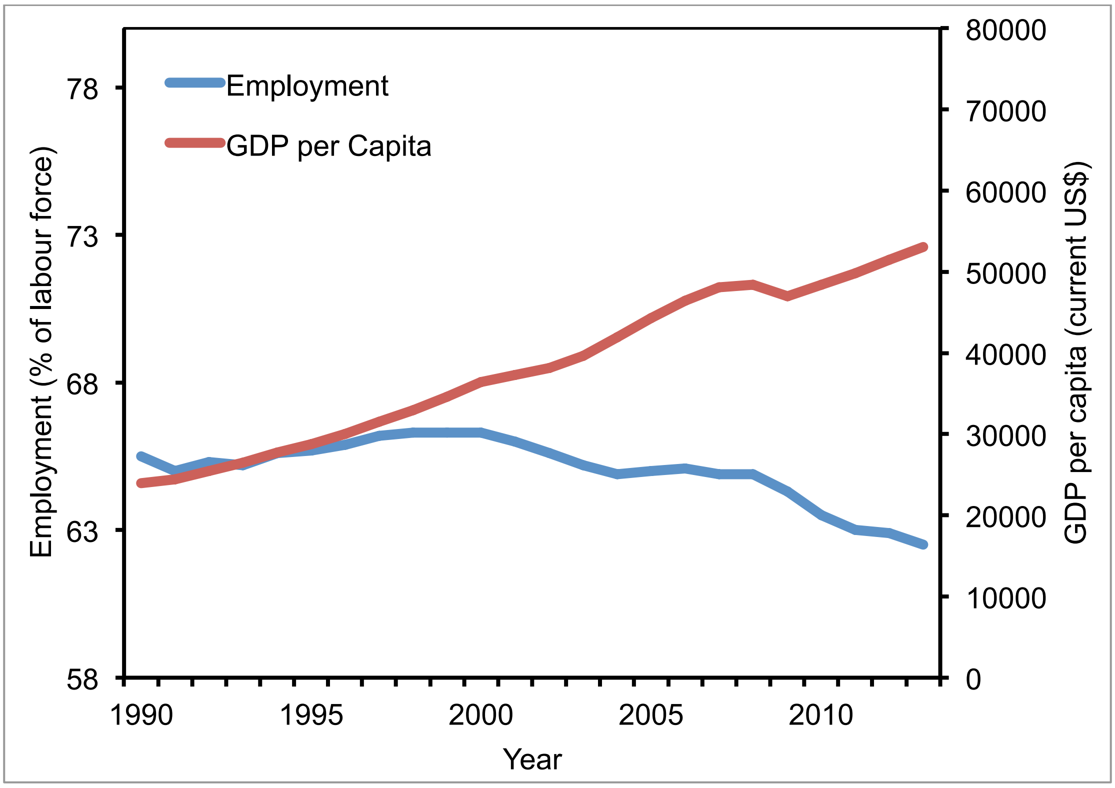

Figure 5.2: Graph showing recent developments in Employment and Economic Growth (The World Bank Databank)

The lack of job growth can have several different explanations but one argument is that technology is replacing labour (Brynjolfsson & McAfee, Race Against the Machine, 2011). Technology is developing fast and it is likely that more jobs will be automated in the future. Frey and Osborne suggest that as many as 47% of American jobs could be at risk of being replaced by machines (Frey & Michael A. Osborne, 2013).

5.3 Job Features

This project aims to investigate the features of different occupations and determine what features affect the automability of a job. The features considered are taken from the O*Net database. This is a database of occupational features expressed “as a standardized, measurable set of variables” (O*Net Resource Center). In total there are 277 different job features listed. The occupational data is collected through surveying randomly selected workers within the desired occupation and the skills and abilities required for the different occupations are “developed by occupational analysts using the updated information from incumbent workers” (O*Net Resource Center). Out of the 277 job features 67 are used in this project. These features were selected because they had data available for all or most of the occupations considered as many occupations do not have data for all features. The data from the database represents the skills and abilities required to perform the occupations and are given a value between 0 and 100, which quantifies the level of a certain skill or ability is needed. The skills used are everything from Gross Body Coordination to Negotiation skills and represent creative, social, mental and physical skills.

It is expected that certain skills and abilities will be particularly important for determining if an occupation can be automated or not. Despite all recent developments in technology there are still limitations in what tasks can be computerised. For example, “computers so far have proved to be great pattern recognizers but lousy general problem solvers”, “have shown little creative ability” and have limited fine motor and complex communication skills (Brynjolfsson & McAfee, Race Against the Machine, 2011). This would imply that abilities such as creative problem solving would be particularly useful for predicting the susceptibility of occupations to computerisation. The paper “The Future of Employment: How Susceptible are Jobs to Computerisation” (Frey & Michael A. Osborne, 2013) uses nine different O*Net features split into three categories to predict the probability of computerisation of over 700 different occupations. The features used belong to the categories Social Intelligence, Creative Intelligence and Perception and Manipulation to reflect the types of tasks where the abilities of computers are still very limited.

6 Gaussian processes

The method used in this project was Gaussian Process Classification. Gaussian Processes (GP) is a non-parametric form of modelling that assumes all variables are normally distributed. An assumed prior together with observed training data forms the posterior distribution over the data. An advantage of using GPs is that it is non-parametric which means that the format of the function defining the relationship between input and output does not need to be specified in advance. The advantage of using Gaussian Processes for this particular application is that it also represents uncertainty in the output. Rather than simply computing the class associated with certain input data the Gaussian process classifier computes the probability that the data point belongs to any particular class. This is useful in this case as it gives the probability that any of the occupations in the experiment can be automated. The classifier trained as part of this experiment used 70 hand labelled data points, where jobs had been labelled as possible to automate ($c = 1$) or impossible to automate ($c = 0$). The labels used were from the paper published by (Frey & Michael A. Osborne, 2013), where a number of academic labelled the occupations they were most certain about. These labels were used as training and test data for all further experiments.

6.1 Covariance function

The Gaussian Process classifier used in this report was implemented in Matlab. An important part of Gaussian Process classification is the covariance function. The covariance function used for most of this project was the exponentiated quadratic covariance function. This function was used as it was found to produce the best classification for the employment data used in the paper by Frey and Osborne. The exponentiated quadratic covariance function is of the form:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \gamma^2 \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{M}(\boldsymbol{x} - \boldsymbol{x}')\right) \quad (6.1)$$

In addition to the exponentiated quadratic another covariance function was used in some experiments for comparison, the Matérn Covariance ($\upsilon = \frac{3}{2}$). The Matérn covariance was not included in the paper by Frey and Osborne and was used in this project for comparison because it is less smooth than the exponentiated quadratic and can also model varying smoothness (Osborne, 2014/15). For the Matérn covariance $\upsilon = \frac{3}{2}$ was used because this is a commonly used version of the kernel and smoother than some other versions, which makes it easier to compare to the exponentiated quadratic. The form of this covariance function is:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \gamma^2(1 + \sqrt{3}|(\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{M}| \exp\left(-\sqrt{3}|(\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{M}|\right) \quad (6.2)$$

The two different covariance functions were compared using the area under the receiver operating characteristics curve (AUC). This measure is equal to 1 for a perfect classifier and 0.5 if the classifier is completely random (Frey & Michael A. Osborne, 2013). Examples of this ROC curve can be seen in Figure 6.1 and Figure 6.2 for both covariance functions.

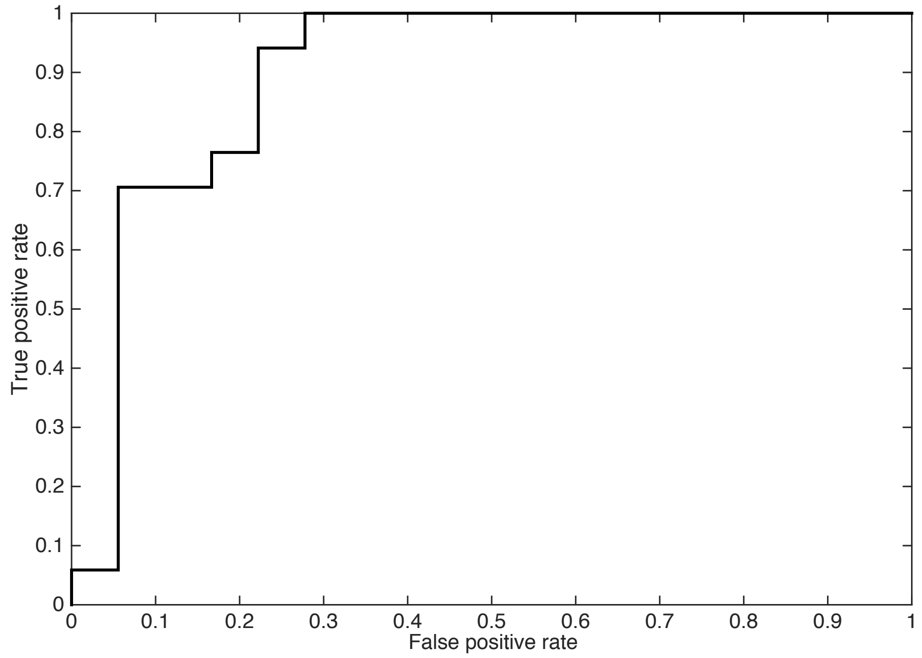

Figure 6.1: ROC for exponentiated Quadratic (AUC=0.8987)

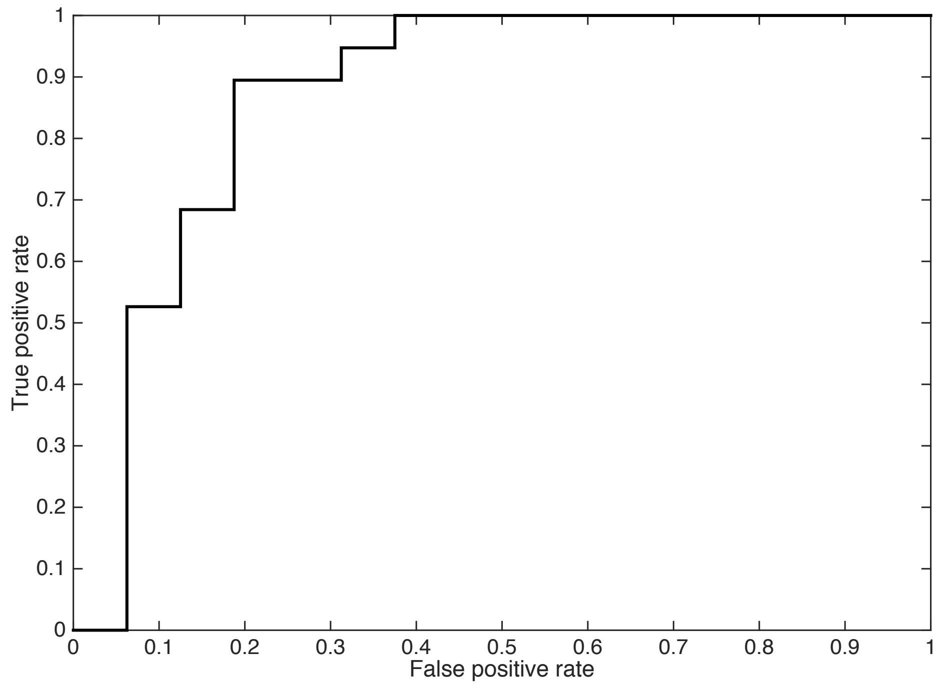

Figure 6.2 ROC for Matérn (AUC=0.8717)

Running the classification 100 times and using 35 randomly selected labelled data points as training data and the other 35 labelled points as test data each time gave the an average AUC for the Matérn covariance 0.8645 and the average AUC for the exponentiated quadratic was 0.8714. This means that the performance of both covariance functions was similar but the exponentiated quadratic Kernel performed the classification slightly better. On the other hand, using marginal likelihood, where a high score indicates good performance, to evaluate classification performance gave the Matérn covariance a higher score of -3052.48 compared to the likelihood when using exponentiated quadratic, which was -5267.46.

6.2 Hyperparameters

Although Gaussian Processes is a non-parametric method of classification there are still hyperparameters that need to be learned. The matrix of hyperparameters ($\boldsymbol{M}$) was defined as diagonal for two reasons. Firstly, it significantly reduces computation time by reducing the number of parameters to be computed from 102 to 12 for a set of 10 features. Secondly, it helps give an understanding of the significance of each feature individually since the value of the associated hyperparameter relates to the importance of that particular feature for the purposes of classification (Murphy, 2012, p. 519-520). The value of the hyperparameters is determined using maximum marginal likelihood. The maximum likelihood was computed in a Matlab script using the optimising function *patternsearch*. This is a direct local optimisation algorithm that searches a pattern of points around the current point specified by a basis matrix and a generating matrix. If any of the points in the pattern improves the objective function this point will be used for the next iteration (Troczon, 1997). One advantage of this optimisation function is that no derivatives are required or computed.

Using the most suitable hyperparameters is important for the classification performance. The hyperparameter associated with each feature is related to the lengthscale. Using the exponentiated quadratic covariance function the matrix $\boldsymbol{M}$ is a diagonal matrix of hyperparameters. These parameters are the square of the inverse lengthscale $\left(\frac{1}{\ell^2}\right)$. The lengthscale associated with a feature determines how much the classification changes with a change in that feature. For example, if the lengthscale of a feature is short this means a small change in the feature can significantly change the output while a longer lengthscale will give a smoother output (Murphy, 2012, p. 519-520). The effect of using different lengthscales can be seen in Figure 6.3, which shows plots of the outputs of one-dimensional Gaussian Process regressions using the same data points but different lengthscales.

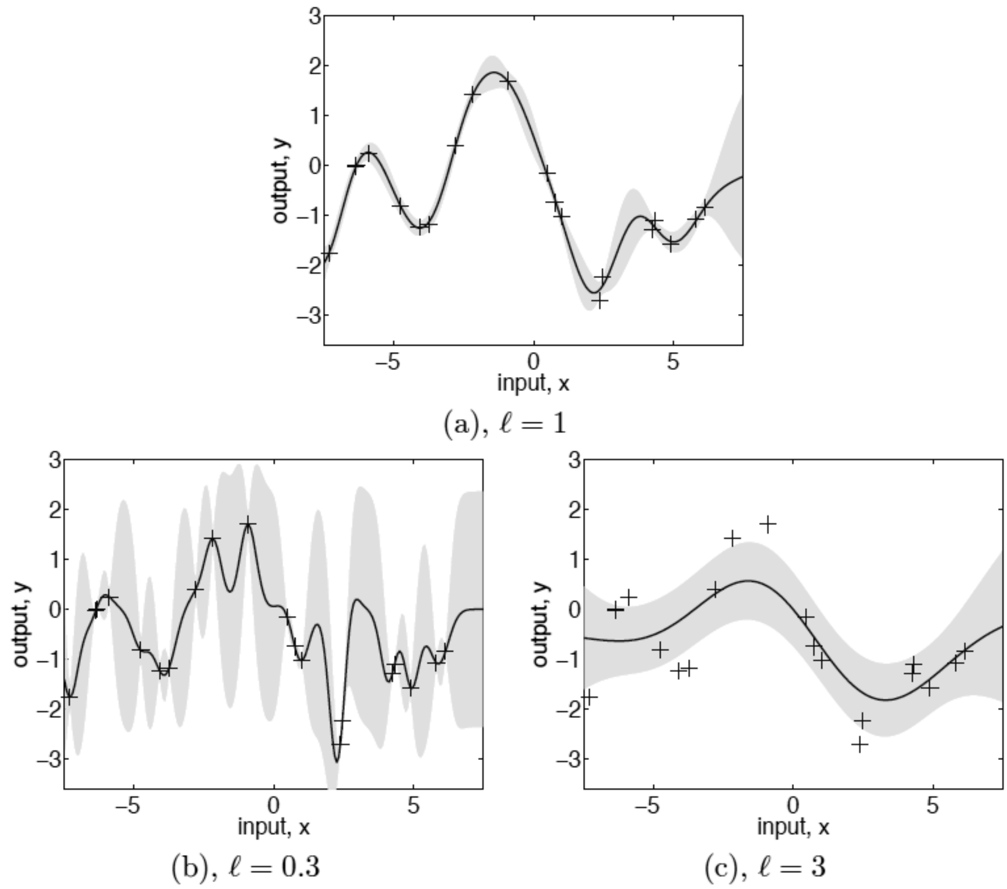

Figure 6.3: One-dimensional Gaussian Process output when using different lengthscales, (a) uses the lengthscale most suitable for the data (b) uses a shorter lengthscale and (c) uses a longer lengthscale. The grey areas show the 95% confidence interval (Rasmussen & Williams, 2006).

6.3 Laplace approximation

In order to perform Gaussian Process Classification some approximations must be made. This is because the mapping between the input data and output class will be non-linear and a Gaussian likelihood function is inappropriate (Rasmussen & Williams, 2006). In order to perform classification a discriminant function y is defined where y can be computed from the input data x. This discriminant function can then be used to compute the probabilistic classification. The mapping from y to class label c is done using the logistic function

$$p(c = 1|y) = \frac{1}{1 + e^{-y}} \tag{6.3}$$

This non-Gaussian distribution is approximated by a Gaussian using Laplace's approximation. This is achieved by fitting an un-normalised Gaussian distribution to the maximum value of the function (Bishop, 2009, p. 213-215). The approximation is found by iterating the function.

$$\boldsymbol{y}^{new} = \boldsymbol{K}_{x,x}\left(\boldsymbol{I} + \boldsymbol{D}\boldsymbol{K}_{x,x}\right)^{-\mathbf{1}}(\boldsymbol{D}\boldsymbol{y} + \boldsymbol{c} - \boldsymbol{\sigma}) \tag{6.4}$$

In the above equation $\boldsymbol{K}$ is the covariance matrix, $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{c}$ is the array of class labels, $\boldsymbol{\sigma}$ is the logistic function array and $\boldsymbol{D}$ is a diagonal noise matrix $D_{ii} = \sigma_i(1 - \sigma_i)$, until it converges at a maximum (Barber, 2012, p. 426), The classification performance could have been improved by using a better approximation such as Expectation Propagation, which is specifically designed for use in Bayesian Networks (Minka, 2001) and is also the approximation method used in (Frey & Michael A. Osborne, 2013). However, this approximation would have been more difficult to implement and since this project is focused on feature selection and relative feature importance as opposed to the absolute probabilities of computerisation it was decided that Laplace approximation would be sufficiently accurate and easier to implement. Using Laplace approximation instead of Expectation propagation does not significantly affect the results. In fact the AUC found in this project is similar to that found by Frey and Osborne in their paper and the only change in methodology is the approximation method. The AUC found when using Laplace approximation was 0.871 and the same figure found by Frey and Osborne when using expectation propagation was 0.894. This shows that expectation propagation gives more accurate classification but using Laplace approximation does not significantly impact the performance.

7 Feature Selection

In order to improve classification performance and better understand what affects whether a job can be automated or not a variety of different feature combinations were investigated. In total 67 features from the O-Net database were used in different combinations. These particular features were selected because they had data available for all or almost all the different occupations used in this project. The focus was on feature selection as opposed to feature learning because this helps understanding the relationships between the features and what makes jobs susceptible to computerisation and this was considered more important than getting the best classification performance.

7.1 Classification Performance Measures

Several different measures were used to determine classification performance. The main measure of performance was marginal likelihood, calculated using

$$\log p(\boldsymbol{c}|\boldsymbol{X}) \approx log\ p(c|\boldsymbol{y}) - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{K}^{-1}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K}| - \frac{1}{2}\log\ |\boldsymbol{K}^{-1} + \boldsymbol{D}| \qquad (7.1)$$

The likelihood was compared to area under the receiver operating characteristic curve (AUC), mean squared error and the mutual information. The values of AUC and mean squared error used were averages from performing the classification 100 times per feature set using different data points as training and test data.

The main measure that was considered the most and was used in the case of conflicting results was marginal log-likelihood. This is because likelihood is a standard figure in Bayesian computations. In addition, it reduces the computation needed compared to AUC or Mean squared error as these methods both require the data to be split into a training and validation set. Splitting the data this way could introduce some bias and therefore the most accurate results are achieved by performing the classification several times using different randomly selected training and validation sets and then computing the average AUC and Mean squared error. This requires more computation than simply calculating the marginal likelihood from the complete set of labelled data. For mutual information and likelihood one computation is enough although the issue with mutual information is that not all the probabilities are known. The formula used to calculate the mutual information is

$$I(C,X) = \sum_{C}\sum_{X} p(x,c)log\left(\frac{p(c,x)}{p(c)p(x)}\right) = \sum_{C}\sum_{X} p(c|x)p(x)log\left(\frac{p(c|x)}{p(c)}\right) \qquad (7.2)$$

In this equation the probability $p\ (c|x)$ is the result of the classification but $p\ (x)$ and $p\ (c)$ are not known. For the purposes of calculating mutual information $p\ (c)$ is assumed to be 0.5 and the empirical distribution is used for $p\ (x)$.

7.2 Feature Performance

In order to achieve the best classification performance and to gain understanding of the significance of the features a variety of feature combinations were used for classification and with varying results. The performance of each of the features individually can be seen in Table **7.1**.

Table 7.1: Classifier performance of different feature sets using different performance measures

Feature	**AUC**	**Mutual information**	**Likelihood**	**Likelihood (Matérn)**
Originality	0.8798	0.2105	-1723.87	-1727.61
Coordination	0.8253	0.1181	-2395.20	-2478.35
Management of Material Resources	0.6550	0.0882	-2403.16	-2634.77
Service Orientation	0.8548	0.1426	-2415.87	-2686.84
Social	0.8470	0.2273	-2494.95	-2714.22

Feature	AUC	Mutual information	Likelihood	Likelihood (Matérn)
Perceptiveness				
Near Vision	0.4564	0.1572	-2614.57	-2884.01
Fluency of Ideas	0.5919	0.1757	-2634.05	-2697.78
Problem Sensitivity	0.4697	0.1272	-3089.22	-3243.50
Systems Evaluation	0.7256	0.0582	-3723.56	-3814.50
Judgment and Decision Making	0.7245	0.1399	-3782.51	-3781.06
Science	0.7001	0.0638	-3871.92	-4.38 E+12
Time Management	0.7697	0.0227	-4048.74	-3707.26
Selective Attention	0.8077	0.0034	-4105.35	-4106.30
Systems Analysis	0.7103	0.0501	-4119.93	-3847.50
Instructing	0.7912	0.0035	-4178.70	-4124.07
Hearing Sensitivity	0.5887	0.0048	-4256.39	-4156.84
Manual dexterity	0.5545	0.0338	-4266.98	-3988.87
Negotiation	0.7052	0.0064	-4289.33	-4138.31
Perceptual Speed	0.7064	0.0047	-4303.43	-4138.96
Speech Clarity	0.7744	0.1223	-4319.89	-3922.89
Time Sharing	0.5961	0.0366	-4372.08	-4044.19
Learning Strategies	0.8377	0.1231	-4399.52	-4026.09
Active Learning	0.6969	0.0592	-4454.96	-4041.02
Persuasion	0.7051	0.0659	-4545.41	-4088.22
Explosive Strength	0.4593	0.0526	-4571.32	-4118.53
Night Vision	0.5511	0.1213	-4593.79	-4087.97
Extent Flexibility	0.5289	0.0462	-4630.67	-376355.18

Feature	AUC	Mutual information	Likelihood	Likelihood (Matérn)
Stamina	0.5252	0.0877	-4630.72	-4144.27
Finger dexterity	0.4239	0.0212	-4632.47	-4160.51
Gross Body Coordination	0.5554	0.0981	-4646.62	-4137.04
Complex Problem Solving	0.7385	0.0620	-4657.75	-4126.50
Speed of Closure	0.5092	0.0232	-4697.05	-4132.94
Static Strength	0.6091	0.0849	-4716.42	-4147.87
Inductive Reasoning	0.8390	0.0982	-4718.58	-4127.97
Auditory Attention	0.5977	0.0001	-4720.54	-4153.62
Depth Perception	0.4831	0.0089	-4736.98	-4161.41
Trunk Strength	0.5492	0.0595	-4809.37	-4195.99
Assisting and caring for others	0.6647	0.1537	-4813.81	-4086.71
Management of Personnel Resources	0.7348	0.0256	-4849.97	-4125.52
Gross Body Equilibrium	0.6070	0.0815	-4855.72	-4163.77
Critical thinking	0.7257	0.0505	-4857.80	-4092.78
Mathematical Reasoning	0.6458	0.0143	-4859.96	-4174.19
Number Facility	0.5890	0.0201	-4931.20	-4176.90
Information Ordering	0.7093	0.0373	-4935.77	-4121.06
Fine arts	0.7126	0.1908	-4941.37	-3972.30
Mathematics	0.5467	0.0378	-4945.51	-4150.89
Deductive Reasoning	0.6287	0.0309	-4961.41	-4113.04

Feature	AUC	Mutual information	Likelihood	Likelihood (Matérn)
Monitoring	0.8309	0.0014	-4970.98	-4178.43
Dynamic Strength	0.4855	0.0502	-5012.78	-4150.58
Flexibility of Closure	0.6688	0.0063	-5019.21	-4156.41
Speaking	0.6709	0.0398	-5037.49	-4121.09
Oral Expression	0.6926	0.0660	-5138.25	-4081.69
Control Precision	0.4178	0.0116	-5144.50	-4180.72
Visualization	0.6781	0.0425	-5224.76	-4176.66
Arm-Hand Steadiness	0.6333	0.0226	-5228.56	-4183.50
Operation and Control	0.5199	0.0456	-5298.87	-4177.13
Category Flexibility	0.5294	0.0551	-5365.34	-4150.22
Active Listening	0.6229	0.0054	-5515.92	-4151.85
Spatial Orientation	0.6394	0.0234	-5538.23	-4129.41
Oral Comprehension	0.5554	0.0810	-5587.19	-4165.51
Writing	0.5887	0.0445	-5649.20	-4162.70
Written Expression	0.5996	0.0723	-5670.07	-4139.54
Cramped work space	0.4213	0.0220	-5712.31	-4204.22
Equipment Maintenance	0.4945	0.0235	-5856.51	-4121.12
Administration and Management	0.5294	0.0371	-5857.89	-4190.47
Reading Comprehension	0.6241	0.0757	-5933.87	-4156.77
Written Comprehension	0.5768	0.0524	-6006.51	-4149.86

7.2.1 Originality

As can be seen from the table Originality is the best performing feature according to both AUC and likelihood using both exponentiated quadratic and Matérn covariance functions. In fact the likelihood is very similar when using the two different covariance functions, particularly for the features where likelihood is high. Originality in this context is defined as "the ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem" (O*Net Online). This fits well with the research by (Bakhshi, Frey, & Osborne, 2015), which shows that creative jobs are at low risk of computerisation. Figure 7.1 shows a graph of a variety of occupations according to their probability of computerisation and creativity. As seen from the graph the creative jobs have a very low probability of computerisation.

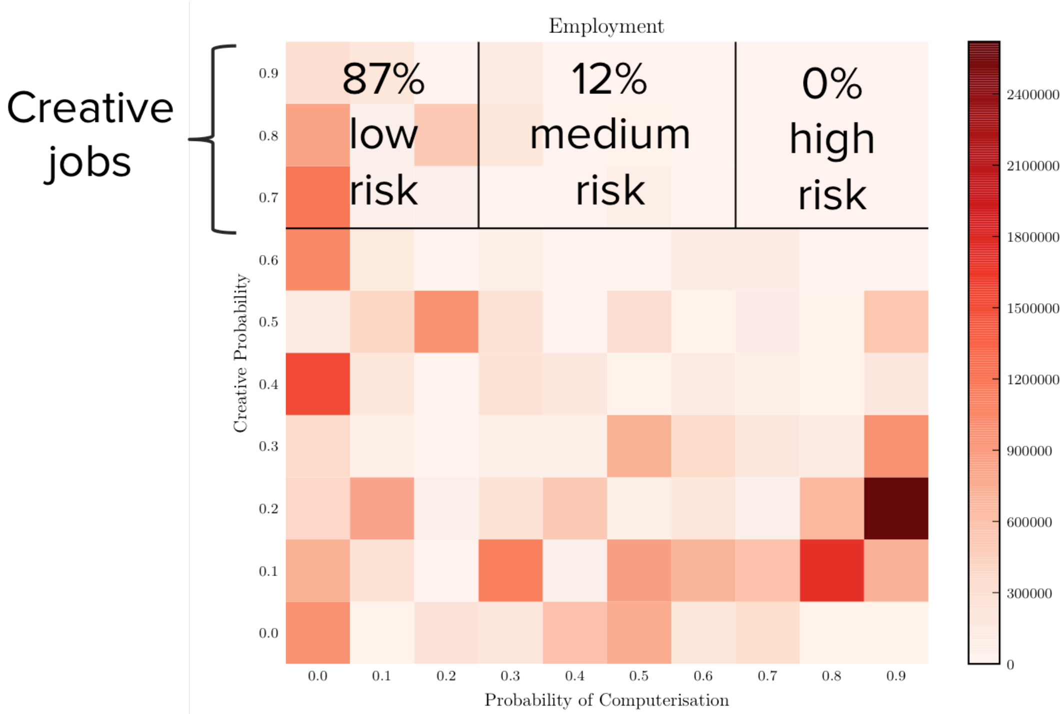

Figure 7.1: Probability of computerisation of jobs of varying creativity (Bakhshi, Frey, & Osborne, 2015)

7.2.2 Social Intelligence and Originality

In addition, several other features perform well according to one or more measures. Furthermore, a feature whose performance is poor individually cannot be assumed to be insignificant as it can still potentially improve performance when combined with another feature. Using more than one feature for classification can result in a better performance than using each of the features individually. For example, Figure 7.2 shows the class labels plotted against Originality and Service Orientation, where Service Orientation is the skill of “actively looking for ways to help people” (O*Net Online). Aside from some outliers the two classes can be clearly separated using a function of the two features and using both features combined results in a better classification than using either of the features separately. As seen in Table **7.1** the likelihood using only Originality is -1723.87 and using only Service Orientation the likelihood is -2415.87. The likelihood when using the two features together is -1721.90, which is higher than both of the individual likelihoods.

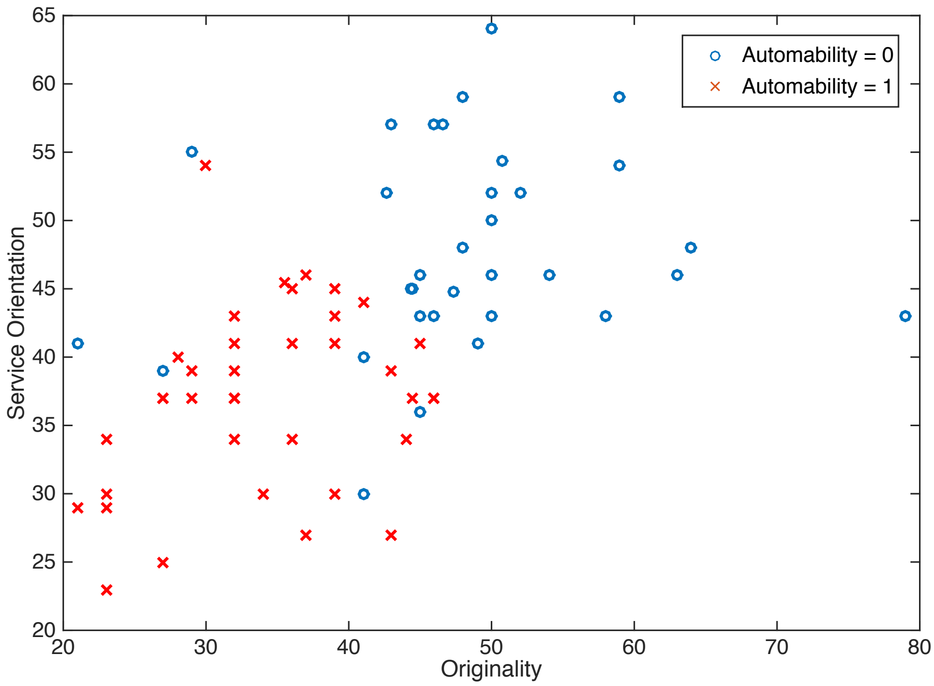

Figure 7.2: Class labels plotted against Service Orientation and Originality

In addition to Service Orientation, the feature Social Perceptiveness seems promising. This feature has the highest figure for Mutual Information and the figures for likelihood and AUC are also relatively high. Like Service Orientation Social Perceptiveness is a feature relating to social skills, which is traditionally difficult to automate (Brynjolfsson & McAfee, Race Against the Machine, 2011). As opposed to Service Orientation combining Social Perceptiveness with Originality does not improve the classification performance. Instead the likelihood when using Social Perceptiveness only is -2494.95 and combining with Originality gives the slightly lower likelihood -2559.11. A possible explanation for this is that Social Perceptiveness could be more closely correlated to Originality. As seen in Figure 7.3 the labels are less spread out than those of Service Orientation combined with Originality. In this plot the labels are collected closer to the diagonal, which could imply some correlation between the two features.

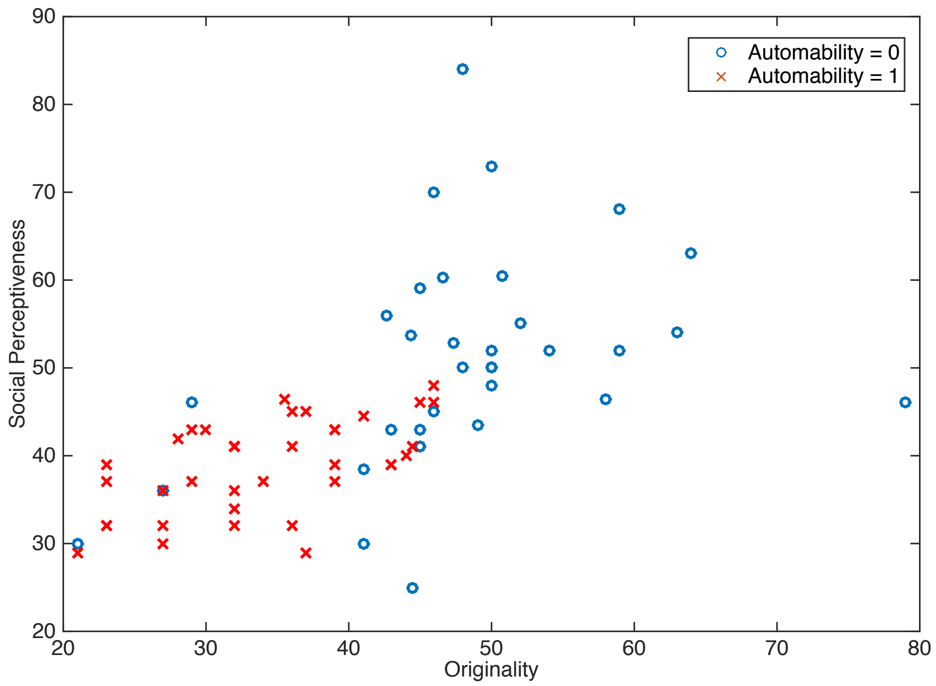

Figure 7.3: Class labels plotted against Social Perceptiveness and Originality

7.2.3 Other Feature Combinations

A variable that seems insignificant by itself but performs well combined with Originality is Time Sharing. As seen from Figure 7.4 it is hard to separate the classes based on Time Sharing only but when combining Time Sharing and Originality there is clearly a cluster of $c = 0$ labels in the top right corner of the plot. This implies that occupation requiring both Time Sharing and Originality are less likely to be automated. In fact the likelihood when using Time Sharing alone is -4372.08 while the likelihood when using the two features combined is -1971.37, which is significantly better than using Time Sharing alone. However, this likelihood is still lower than that of only Originality.

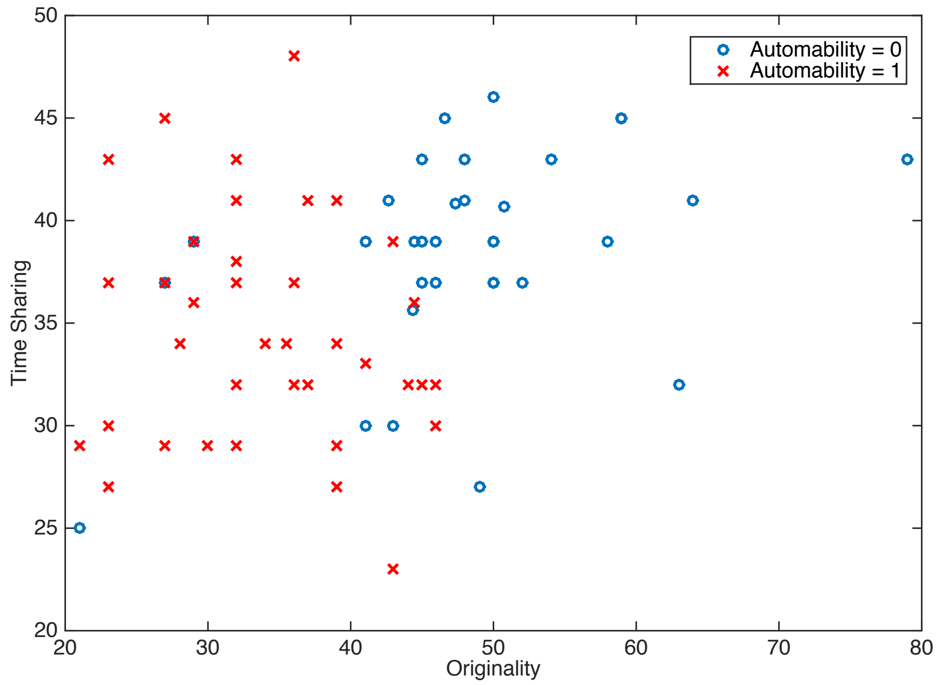

Figure 7.4: Class labels plotted against Time Sharing and Originality

Another feature that performs significantly better when combined with Originality is Gross Body Coordination. This is an interesting feature combination because using both features together makes the likelihood lower than each feature separately but the AUC is significantly improved. Figure 7.5 shows that Gross Body Coordination alone is not a good predictor of automability but combining the features generates a cluster of $c = 1$ labels in the bottom left corner of the plot meaning that occupations requiring low Gross Body Coordination and Originality have a high probability of automation. The likelihood when using Gross Body Coordination alone is -4646.62 and combined with Originality the likelihood is -5847.16. This means that based on likelihood Gross Body Coordination gives poor classification when combined with Originality. However, using the two features combined gives a significant improvement in AUC. The AUC when using Originality only is 0.8798, the AUC using Gross body Coordination is 0.555, which is a very poor figure, but combining the two features gives an AUC as high as 0.942. This example shows that the different performance measures can differ significantly, which emphasises the importance of selecting the right measure.

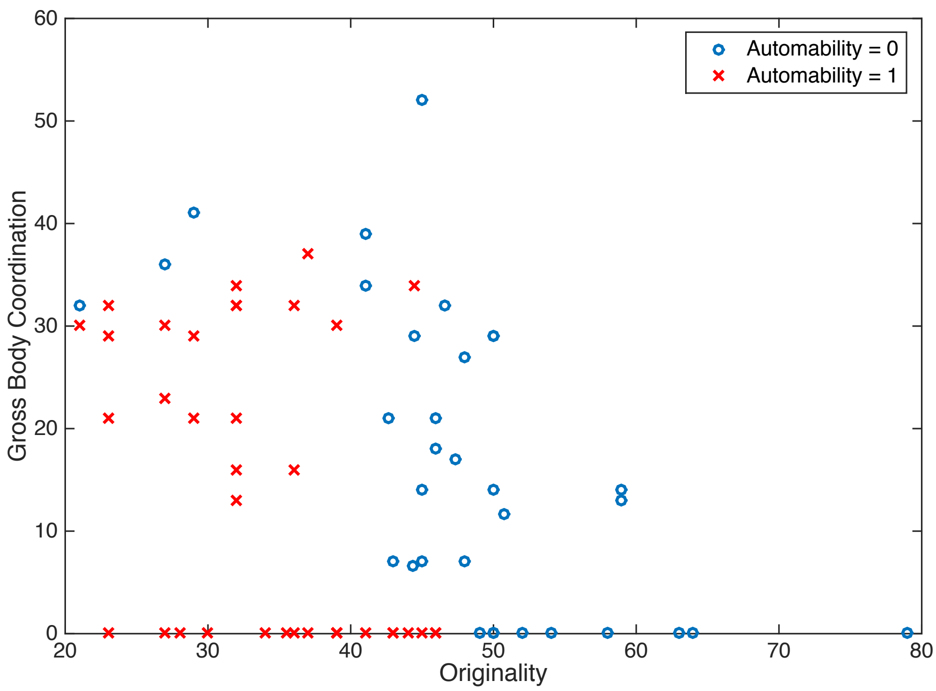

Figure 7.5: Class labels plotted against Gross Body Coordination and Originality

7.2.4 Predicted probabilities

The feature set used for classification determines the probability of computerisation predicted by the model for each occupation. Figure 7.6 shows a plot of probability of automation as a function of Originality for different feature sets. The result of the set of nine features used by Frey and Osborne is included in the plot as well as the result of using only Originality and the training class labels. The probabilities when based on nine features are more spread out than those based on only originality, which form a smooth curve. This is expected since the former is a function of multiple variables while the latter is a function of originality only. It can also be seen that the probabilities based on originality only reach lower minimum probabilities and higher maximum probabilities thereby being closer to some of the class labels. However, these probabilities cannot make predictions for where class labels overlap around Originality=40. It can also be seen that around the ends where there are no training points the probability tends to 0.5, which is the prior probability.

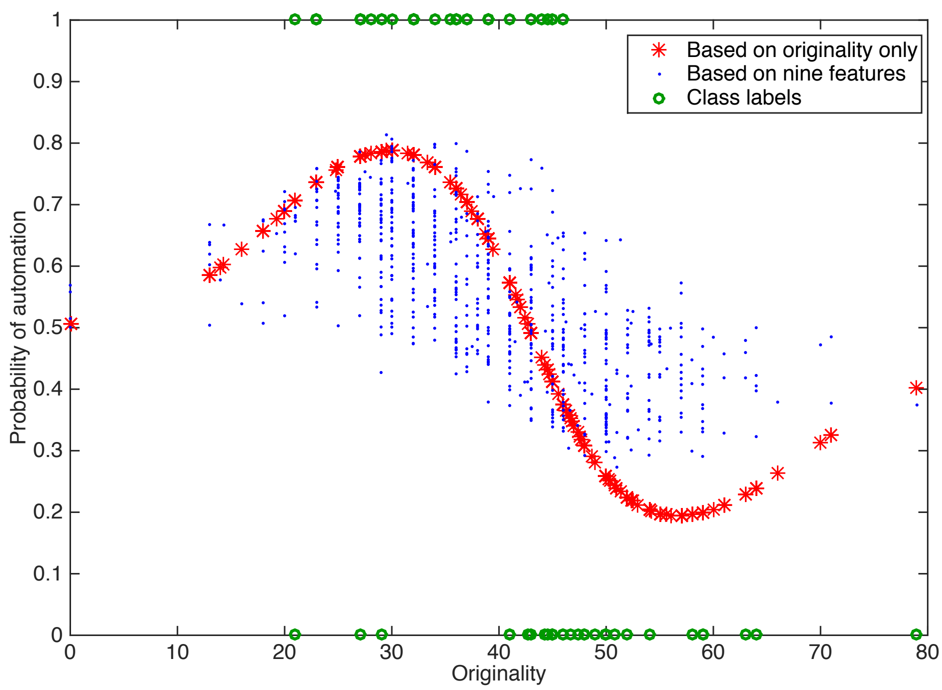

Figure 7.6: Probability of computerisation as function of Originality based on different feature sets

The best classification performance found so far was achieved using a feature set consisting of Originality and Service Orientation. The resulting probabilities of computerisation from this feature set are included in Figure 7.7. These probabilities are less centred around 0.5 than those based on nine features and the maximum probabilities are even higher than those based on originality only and the minimums are lower.

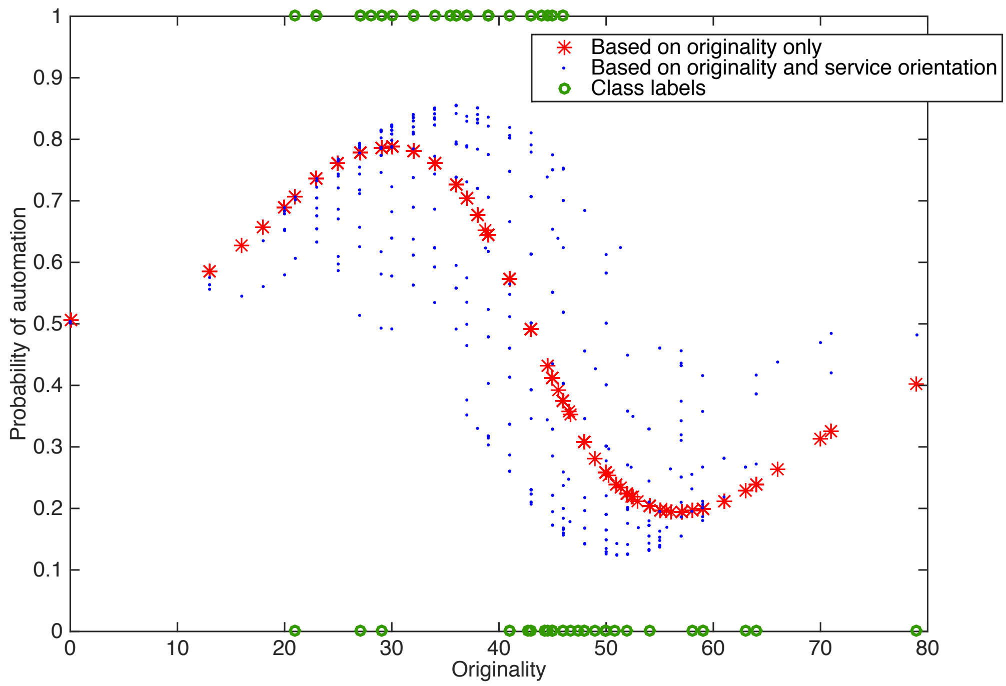

Figure 7.7: Probability of computerisation as function of Originality using different feature sets for classification

When Originality and Service Orientation are combined the likelihoods imply that the more significant feature is Originality. In Figure 7.8 the difference between using only Service Orientation and combining it with Originality can be seen.

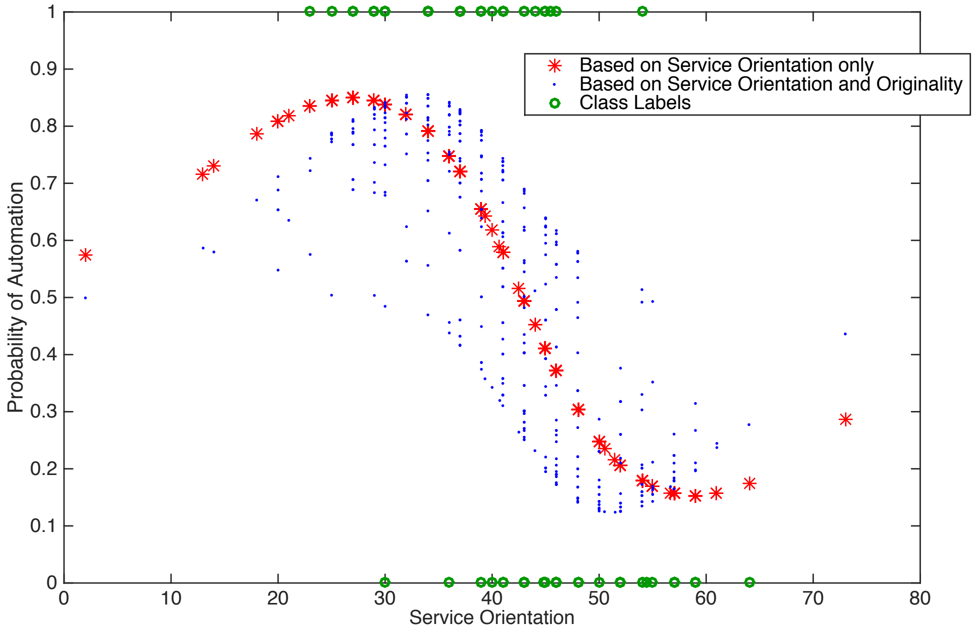

Figure 7.8: Probability of computerisation as function of Service Orientation using different feature sets for classification

The probabilities when using only Service Orientation form a smooth curve as expected. The probabilities found using both Service Orientation and Originality also vary relatively smoothly with Service Orientation although the curve is slightly shifted. Looking at the two graphs it is not clear that Originality is more significant when the two variables are combined. However, the hyperparameters associated with each of the variables can also help determine which feature is given more weight when computing probability of computerisation as they are related to the lengthscales. When using a feature set consisting of Originality and Service Orientation only the hyperparameter associated with Originality is 0.0182 and the hyperparameter associated with Service Orientation is 0.0086. This means that the lengthscale of Originality is 7.41 and the lengthscale of Service Orientation is 10.75 meaning that a change in Originality will have a larger impact on the impact but the lengthscales are similar meaning that both features are significant when computing the classification.

Changing the feature set used for classification will change the resulting probabilities of computerisation. There is a change in probability of computerisation of each occupation when using two features compared to using only one. Changing the feature set from Originality to Originality combined with Service Orientation results in an average absolute change in probability of computerisation of 0.1111 and a standard deviation of 0.0936. The equivalent average change in probability when going from a feature set of Service Orientation only to the combined set is 0.1091 and the standard deviation is 0.0819. The biggest change in probability seen when introducing Originality in addition to Service Orientation is that of Mathematicians. When using Service Orientation only the predicted probability of computerisation is as high as 0.839. This is because the level of Service Orientation required is only 30. However, once Originality, where the level required is 71, is introduced the probability falls to 0.485. It is likely that the true probability could be even lower than this. When using Originality only the predicted probability is 0.325. The probability falls close to 0.5 when using both features because the features are in the region where training data is sparse meaning that the posterior probability approaches 0.5. A table with various probabilities predicted using a feature set consisting of Originality and Service Orientation can be seen in Appendix A.

7.3 Greedy Selection

In order to find the best combination of features, a greedy forward selection, as outlined in "An Introduction to Variable and Feature Selection" (Guyon & Elisseeff, 2003), was performed. This meant that each feature individually was used to perform classification and the feature that performed best was selected to be part of the set. The classification was then repeated with the newly selected feature combined with each of the remaining features in turn and this was used as the feature set for classification. Again the feature combination that resulted in the best classification was selected. Greedy selection was used because it is a relatively fast method for finding good feature sets and it is also a good method for preventing overfitting (Reunanen, 2003). The disadvantage of using Greedy forward selection is that not all combinations of features are explored and once a feature is selected it will never be removed from the set even if this would improve performance (Guyon & Elisseeff, 2003). The performance measure used to select the features generating the best classification was marginal log-likelihood. When an iteration is performed the feature that results the highest marginal likelihood as calculated using Equation 7.1 is added to the set.

7.3.1 Best Feature Combination

The individually best feature based on likelihood is Originality. This feature was selected to be part of the feature set. Adding a further feature to the set and evaluating performance shows that the best classification based on likelihood is achieved by adding Service Orientation to the set already consisting of Originality. The likelihood when using these two features is -1721.90, which is better than either of the features individually. A heat map of the classification based on these features can be seen in Figure 7.9. From the plot it can be seen that the classifier performs well in the area around the training data. Further away the probability of automation converges to 0.5. This happens because the prior probability was set to 0.5, meaning that without any additional information from training data this will be the posterior probability predicted by the model. The model also predicts a probability of automation in the area between the two sets of labels. Since there are only two variables used to predict the probability of automation there is a region where the labels overlap. It is possible that taking further features into account could help determine the automability of occupations that fall in this region. Appendix A contains a table showing the predicted probability of computerisation of various occupations.

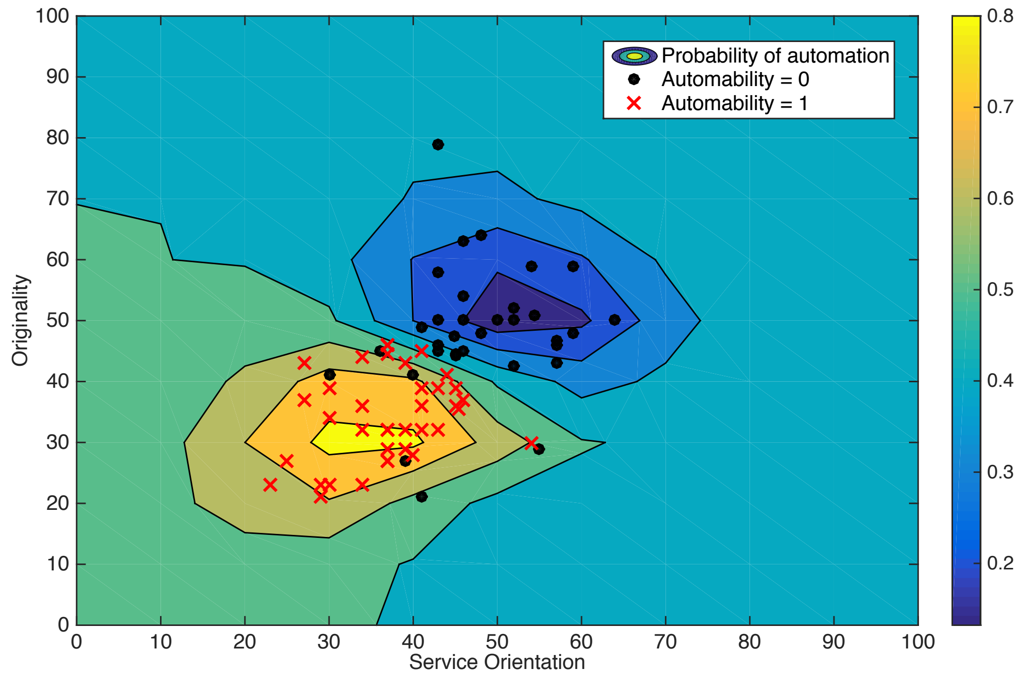

Figure 7.9: Heat map showing probability of computerisation as function of Originality and Service Orientation

7.3.2 Adding Further Features

Adding more features than this does in fact not lead to better performance measured in terms of likelihood. Running the classification for more and more features, adding the one that gives the best marginal likelihood every time achieves the likelihoods seen in Table **7.2**.

Table 7.2: Results of greedy selection algorithm

Number of features	**Feature Added**	**Maximum likelihood**
1	Originality	-1723.87
2	Service Orientation	-1721.90
3	Inductive Reasoning	-1875.75
4	Writing	-2817.27
5	Time Sharing	-1779.83
6	Social Perceptiveness	-1881.11
7	Control Precision	-2057.53
8	Learning Strategies	-1935.29
9	Written Expression	-1909.22
10	Judgment and Decision Making	-2243.61
11	Flexibility of Closure	-2667.33
12	Auditory Attention	-2456.38
13	Fine Arts	-2473.55
14	Trunk Strength	-3462.66
15	Arm-Hand Steadiness	-3556.24
16	Active Listening	-4202.30
17	Cramped Work Space	-4441.95
18	Extent Flexibility	-4299.92
19	Mathematics	-4312.71
20	Operation and Control	-5058.24

As can be seen in the above table the maximum likelihood is achieved with two features. This data can be seen in a plot of likelihood as a function of number of features in Figure

7.10. As seen from the graph the likelihood gets lower when adding more features. One explanation for this is the fact that there is probably some redundant information when using several different features, as different features will be correlated. Furthermore, since there are only 70 training data point it becomes difficult for the model to learn the relationship between large numbers of variables. Using all variables would give a set of 67 features and 70 training points would be far from sufficient to model the effect of each of those features.

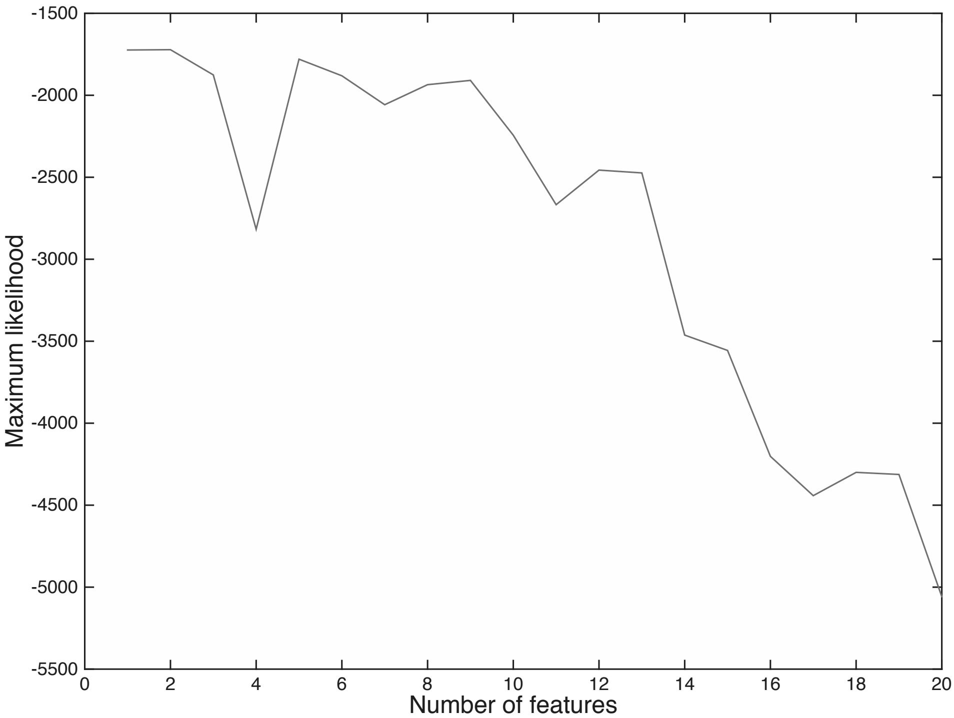

Figure 7.10: Maximum likelihood achieved as function of number of features in feature set

Using a measure of performance other than likelihood gives completely different results. Based on mutual information Social Perceptiveness is the individually best feature and further feature selection would also give a very different result.

7.4 Alternative Feature Selection Methods

In addition to Greedy Selection some other feature selection methods have been explored in order to find the optimum feature set.

7.4.1 Variable Ranking

Feature selection using variable ranking involves performing the classification using each of the variables separately and then selecting the most promising ones to make up the feature set (Guyon & Elisseeff, 2003). The advantage of this method is that it is very simple and easy to implement and it is useful for finding the features that are individually the most significant. However, this method does not take into account any correlation between the features or how well particular features perform together. The set of two features selected using variable ranking will be the two features that achieve the highest marginal likelihood individually. As seen in Table **7.1** the features selected in this experiment would be Originality and Coordination. The marginal likelihood achieved using this feature set is -2419.47, which shows that this method does not find the optimum solution.

7.4.2 Exhaustive Search

An exhaustive search was performed over all possible sets of two features. This involved running the classification for every possible feature set and comparing the marginal likelihood achieved in each classification. This method is certain to find the optimum set of two features but it is also very computationally heavy. The exhaustive search confirmed that a feature set consisting of Originality and Service Orientation is in fact the optimum set of two features as no other combination of two features could generate a higher marginal likelihood.

An exhaustive search over more than two variables was not performed as adding more variables would significantly increase the number on classifications to be performed. There are 4422 (67×66) different combinations of two features and adding only one more feature would increase this number to 287 430 ($67 \times 66 \times 65$).

7.5 Linear Embedding

In an attempt to further improve classification performance linear embedding methods were explored. This is an experiment to explore feature learning as opposed to feature selection, which has been the main focus in this project. Feature learning can sometimes achieve better performance than feature selection, as an ideal feature set could be a linear combination of several features. This type of dimensionality reduction has the advantages of compressing the data and extracting the relevant information from a large set of features, which can improve classification performance (Mohri, Rostamizadeh, & Talwalkar, 2012). To achieve this feature learning a linear embedding was used. The set of 10 features producing the best likelihood based on the greedy selection were used and were embedded to a set of 3 features that were linear combinations of the ten features. The covariance function used was

$$k(\boldsymbol{x}, \boldsymbol{x}') = \gamma^2 \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')\boldsymbol{R}^T\boldsymbol{R}(\boldsymbol{x} - \boldsymbol{x}')^T\right] \tag{7.3}$$

In the above equation $\boldsymbol{R}$ is the embedding matrix and the parameters of this matrix were found by using maximum marginal likelihood, which is computed using Equation 7.1.

7.5.1 Results

The maximum likelihood was achieved with the following parameters

$$\gamma = 1.0604$$

$$\boldsymbol{R} = \begin{bmatrix} 1.8120 & -0.6584 & 0.8541 \\ 1.2989 & -0.0252 & 0.4955 \\ -2.0591 & -1.6791 & -0.3821 \\ 1.4718 & -3.5906 & 1.9819 \\ -0.3053 & -0.5277 & 1.1277 \\ 1.5697 & -0.2768 & 0.1144 \\ 1.7433 & -0.3416 & -0.4913 \\ 0.0065 & 1.4848 & 0.8558 \\ 1.3222 & 0.6795 & 5.4282 \\ 1.8962 & 1.2370 & 0.3570 \end{bmatrix}^T \tag{7.4}$$

The performance of this classification was compared to that of random embeddings (Wang, Zoghi, Matheson, Hutter, & de Freitas, 2013). The parameters of $\boldsymbol{R}$ were all randomly generated and only γ was optimised using maximum likelihood. The likelihood found was the average likelihood achieved by performing 100 random embeddings of the same format as previously. It turned out to be very difficult to compute the optimum R because the likelihood function was highly multimodal. In fact over 150 unique local maximum points have been found when trying to find the global maximum. In order to compute the global optimum R the Matlab function *Multistart* was used. This is a function that performs local optimization at multiple uniformly distributed starting points (MathWorks). The local optimiser used was *Fmincon*. The best likelihood computed using linear embeddings where the parameters were optimised using maximum likelihood was -4112.08. This can be compared to the average likelihood from running random embeddings of the same format, which was -4116.3. The likelihood achieved when optimising the embedding was surprisingly low given that combinations of the features had previously generated better performance.

There are several possible explanations of why this is the case. Firstly, as mentioned the likelihood function is highly multimodal making it difficult to find the global maximum.

Furthermore, an optimising method that relies on running local optimisation several times would need to be run many times to generate an accurate optimum value. Because the target function is 31-dimensional it would need to be evaluated 2^{31} times just to have tested two different starting points in every dimension. Due to time restriction and computation time the optimisation was never performed for more than 1000 starting points. It might be the case that the optimisation method used in this case was not the most suitable. Rather than repeating local optimisations it could be that a global optimiser such as DIRECT "Lipschitzian Optimization without the Lipschitz Constant", which searches the entire function for the global optimum point (Jones, Perttunen, & Stuckman, 1993). Finally, because there were only 70 training data points it will be difficult to accurately compute 31 different parameters based on the training data.

7.5.2 Embedding a Smaller Feature Set

In order to overcome some of the problems arising from having many parameters to compute an alternative embedding was attempted by linearly combining 4 features into 2 features through embedding. This reduced the number of parameters to compute from 31 to 9. This time the 4 first variables selected using the greedy selection were used. This resulted in a slightly improved likelihood of -4073.17, which was achieved with the following parameters

$$\gamma = 1.0612$$

$$\boldsymbol{R} = \begin{bmatrix} 2.7552 & -2.0811 \\ -1.4821 & 2.9428 \\ 0.2802 & 1.3090 \\ -0.5134 & 1.3195 \end{bmatrix}^T \qquad (7.5)$$

This embedding means that the two learned features are:

$$\begin{aligned} x_1 &= 2.7552 \times \text{Originality} - 1.4821 \times \text{Service Orientation} \\ &+0.2802 \times \text{Inductive Reasoning} - 0.5134 \times \text{Writing} \end{aligned} \tag{7.6}$$

$$\begin{aligned} x_2 &= -2.0811 \times \text{Originality} + 2.9428 \times \text{Service Orientation} \\ &+1.3090 \times \text{Inductive Reasoning} + 1.3195 \times \text{Writing} \end{aligned} \tag{7.7}$$

This is a small improvement in overall likelihood but a significant improvement compared to using a random embedding. The average likelihood when using 100 randomly generated embeddings is -87670.61, which is a much lower likelihood that that achieved by optimising the embedding parameters. It was also noted that the likelihood from each random embedding differed more than in the case with 30 parameters of $\boldsymbol{R}$ where the likelihoods achieved were all very similar. The likelihood produced when using randomly generated embedding from 4 to 2 features varies from -4115.95 to -7249795.07.

It is possible that there could exist better solutions to this problem but this experiment has failed to show that a weighted combination of the features can generate better classification performance than using the features individually.

8 Conclusions

In summary, it seems Originality is the single most important feature to determine if a job is susceptible to computerisation. This means that jobs requiring significant creative problem solving are unlikely to be computerised in the near future while jobs requiring less original thinking will be at higher risk of computerisation. This confirms the argument by Brynjolfsson and McAfee (Brynjolfsson & McAfee, Race Against the Machine, 2011) that computers are still very limited in the ability of general problem solving. The fact that originality is the single most important feature for predicting computerisation suggests that this is the ability where computers are most limited. Another reason why originality is more significant than many other variables could be the fact that originality varies greatly between different occupations. The level of originality required varies evenly from 0 to 79. Other variables such as fine arts will be less significant since more than half of the occupations have a value of 0.

The best feature set found was Originality combined with Service Orientation. The combination of these two features gave better performance than any other combination of features explored as part of this analysis. This implies that the combinations of tasks that are most difficult to automate are those that require creative problem solving and actively looking to help people. This agrees with previous research that identifies creativity and social intelligence as areas where computers are still very limited. It is possible, however, that taking more features into account will be better in some cases as there are many more features that make up an occupation. This is simply the combination of features that gave the best overall performance when applied to the 70 training data points used in this experiment.

8.1 Limitations

There are some limitations with the analysis performed in this project. The data labels used for training and testing the model were all labelled by hand. This means there could be errors in the labels and given that there are only a total of 70 training data points this could affect the results. Ideally more training data would be desirable. For example, when doing experiments with embedding 70 training points are used to learn 31 parameters. This is not enough training data to properly learn the significance of all parameters, which can partly explain the poor results from the embedding experiments.

Furthermore, Laplace approximation is not necessarily the most appropriate method for approximating the non-Gaussian part of the model. As mentioned in section 6.3 a method such as Expectation Propagation would give more accurate results.

Finally, this analysis attempts only to explore which jobs can be automated. As David Autor points out “even when a task is fully codified, however, this does not mean it *will* be automated” but it depends on relative costs and other factors (Autor, The "task approach" to labour markets: an overview, 2013). This project considers only whether complete occupations can be performed by a machine and does not consider individual tasks that might be replaced. For some occupations it might make sense to replace certain tasks with machines but not the job as a whole. In addition, this analysis does not take into account the fact that humans will be preferred for certain jobs regardless if a machine could perform the same tasks. One such occupation is professional athletes, which according to (Frey & Michael A. Osborne, 2013) has a probability of computerisation of 28%. This probability implies that there is a chance robots could perform all the tasks of athletes, however, it is unlikely that machines will actually replace professional sports competitors.

9 References

Autor, D. H. (2013). The "task approach" to labour markets: an overview. *Journal for Labour Market Research* .

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *Quarterly Journal of Economics* .

Bakhshi, H., Frey, C. B., & Osborne, M. (2015). *Creativity vs Robots: The Creative Economy and the Future of Employment.* Nesta.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning.* Cambridge: Cambridge University Press.

Bishop, C. M. (2009). *Pattern Recognition and Machine Learning* (8th ed.). Springer.

Brynjolfsson, E., & McAfee, A. (2011). *Race Against the Machine.* Lexington, Massachusetts: Digital Frontier Press.

Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age.* W. W. Norton & Company.

Frey, C. B., & Michael A. Osborne. (2013). The Future of Employment: How Susceptible are Jobs to Computerisation.

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research , 3*, 1157-1182.

Jaimovich, N., & Siu, H. E. (2012). The Trend is the Cycle: Job Polarization and Jobless Recoveries. *NBER Working Paper No. 18334* .

Jones, D. R., Perttunen, C. D., & Stuckman, B. E. (1993). Lipschitzian Optimization Without the Lipschitz Constant. *Journal of Optimization Theory and Application , 99*, 157-181.

Keynes, J. M. (1930). *Economic Possibilities for our Grandchildren.*

MathWorks. (n.d.). *How GlobalSearch and MultiStart Work*. Retrieved January 2015 from MathWorks Documentation: http://uk.mathworks.com/help/gads/how-globalsearch-and-multistart-work.html

Minka, T. P. (2001). Expectation Propagation for Approximate Bayesian Inference. *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence* (pp. 362-369). Morgan Kaufmann Publishers Inc.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of Machine Learning.* Cambridge, Massachusetts: The MIT Press.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective.* Cambridge, Massachusetts: The MIT Press.

O*Net Online. (n.d.). *Browse by O*NET Data*. Retrieved January 2015 from O*Net Online: http://www.onetonline.org/find/descriptor/browse

O*Net Resource Center. (n.d.). *About O*Net*. Retrieved February 2015 from O*Net Resource Center: http://www.onetcenter.org/overview.html

O*Net Resource Center. (n.d.). *O*NET® Data Collection Overview*. Retrieved January 2015 from O*Net Resource Center: http://www.onetcenter.org/dataCollection.html

Osborne, M. A. (2014/15). *C24 advanced probability theory lecture notes.* From http://www.robots.ox.ac.uk/~mosb/teaching/C24/2_slides_Gaussian_processes.pdf

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian Processes for Machine Learning.* Cambridge, Massachusetts: The MIT Press.

Reunanen, J. (2003). Overfitting in Making Comparisons Between Variable Selection Methods. *Journal of Machine Learning Research* , 1371-1382.

The Economist. (2014, January 18). The Future of Jobs: The onrushing wave. *The Economist* .

The World Bank Databank. (n.d.). *Databank*. Retrieved March 2015 from The World Bank: http://data.worldbank.org/indicator/SL.TLF.CACT.ZS/countries/US?display=graph

Troczon, V. (1997). On the Convergence of Pattern Search Algorithms. *SIAM J. Optim. , 7*, 1-25.

Wang, Z., Zoghi, M., Matheson, D., Hutter, F., & de Freitas, N. (2013). Bayesian Optimization in a Billion Dimensions via Random Embeddings. *arXiv preprint arXiv:1301.1942* .

10 Appendix A: Probabilities of Computerisation

Occupation	**Originality**	**Service Orientation**	**Label**	**Probability**
Special Education Teachers, Secondary School	52	52		0.1254
Lawyers	52	52	0	01254
Audiologists	54	52		0.1331
Elementary School Teachers, Except Special Education	55	50		0.1373
Air Traffic Controllers	52	48		0.1413
Pharmacists	48	54		0.1424
Travel Guides	48	50		0.1429
Medical Scientists, Except Epidemiologists	55	48		0.1479
Human Resources Managers	55	54		0.1531
Education Administrators, Postsecondary	55	54		0.1531
Marketing Managers	57	50		0.1550
Chiropractors	46	54		0.1599
Lodging Managers	50	57		0.1649
Advertising Sales Agents	46	55		0.1658
Chefs and Head Cooks	48	48	0	0.1679
Database Administrators	54	46		0.1725
Landscape Architects	54	46	0	0.1725
Physical Therapists	45	54		0.1728
Veterinarians	45	54		0.1728
Registered Nurses	46.625	57	0	0.1785
Public Relations Specialists	50	46		0.1902

Occupation	Originality	Service Orientation	Label	Probability
Substance Abuse and Behavioral Disorder Counselors	46	57	0	0.1839
Epidemiologists	50	46		0.1902
Environmental Engineers	55	45		0.1950
Sociologists	55	45		0.1950
Computer Systems Analysts	58	46		0.1952
Marriage and Family Therapists	48	59	0	0.1964
Mental Health Counselors	57	57		0.2099
Librarians	43	52		0.2102
Floral Designers	50	45		0.2206
Fitness Trainers and Aerobics Instructors	45	48		0.2212
Concierges	43	57	0	0.2234
Travel Agents	46	61		0.2372
Choreographers	57	43		0.2510
Astronomers	57	43		0.2510
Civil Engineers	58	43	0	0.2554
Music Directors and Composers	56	42.5		0.2642
Fashion Designers	63	46	0	0.2672
Photographers	52	43		0.2707
Chief Executives	64	48	0	0.2722
Clergy	50	64	0	0.2774
Mechanical Engineers	61	43		0.2815
Industrial Production Managers	50.25	43		0.2967
Animal Scientists	50	43		0.3013

Occupation	Originality	Service Orientation	Label	Probability
Construction Managers	50	43		0.3013
Radio and Television Announcers	50	43		0.3013
Geographers	50	43		0.3013
Electrical Engineers	50	43	0	0.3013
Retail Salespersons	39	54		0.3032
Materials Scientists	57	41		0.3105
Insurance Sales Agents	39	52		0.3177
Engineers, All Other	57	40.7143		0.3196
Aerospace Engineers	54	41		0.3291
Political Scientists	54	41		0.3291
Skincare Specialists	41	48		0.3368
Childcare Workers	44.5	45		0.3441
Real Estate Sales Agents	37	52		0.3762
Microbiologists	64	39		0.3862
Chemical Engineers	64	39		0.3862
Zoologists and Wildlife Biologists	46	43	0	0.4031
Biomedical Engineers	71	43		0.4203
Economists	49	41	0	0.4271
Biochemists and Biophysicists	66	36		0.4379
Financial Analysts	48	41		0.4559
Computer Programmers	48	41		0.4559
Interpreters and Translators	41	45		0.4605
Editors	55	37		0.4607
Physicists	79	43	0	0.4820

Occupation	Originality	Service Orientation	Label	Probability
Mathematicians	71	30		0.4845
Loan Officers	30	54	1	0.4918
Flight Attendants	29	55	0	0.4930
Dancers	50	39		0.5011
Historians	0	29		0.5037
Accountants and Auditors	41	44	1	0.5117
Electricians	46	41		0.5189
Credit Analysts	39	45	1	0.5234
Insurance Underwriters	37	46	1	0.5351
Funeral Attendants	16	43		0.5450
Judicial Law Clerks	45	41	1	0.5513
Teacher Assistants	41	43		0.5647
Massage Therapists	37	45		0.5748
Film and Video Editors	50	37		0.5827
Telemarketers	34	46		0.5924
Maids and Housekeeping Cleaners	21	41	0	0.6065
Dental Hygienists	32	46		0.6117
Reporters and Correspondents	50	36		0.6127
Correspondence Clerks	39	43		0.6177
Statisticians	51.3333	32		0.6240
Postal Service Clerks	27	45		0.6254
Slaughterers and Meat Packers	18	21		0.6351
Medical Secretaries	30	43		0.6822
Biological Technicians	48	30		0.6843

Occupation	Originality	Service Orientation	Label	Probability
Gaming Dealers	32	43	1	0.6897
Cost Estimators	46	37	1	0.7005
Market Research Analysts and Marketing Specialists	46	37	1	0.7005
Dishwashers	21	29	1	0.7016
Light Truck or Delivery Services Drivers	23	34	1	0.7047
Paralegals and Legal Assistants	39	41	1	0.7058
Technical Writers	43	27	1	0.7067
Civil Engineering Technicians	43	39	1	0.7069
Waiters an d Waitresses	27	39	0	0.7177
Medical Transcriptionists	23	23	1	0.7221
Food Preparation Workers	23	30		0.7324
Bicycle Repairers	32	41		0.7378
Taxi Drivers and Chauffeurs	32	41	1	0.7378
Motorboat Operators	36	41	1	0.7383
Surveyors	44.5	37	1	0.7387
Bus Drivers, Transit and Intercity	27	37	1	0.7436
Cashiers	29	39	1	0.7461
Commercial Pilots	41	39		0.7478
Athletes and Sports Competitors	45	36	0	0.7505
Actors	46	34		0.7531
Farm Labor Contractors	29	37	1	0.7738
Security Guards	29	37		0.7738

Occupation	Originality	Service Orientation	Label	Probability
Industrial Truck and Tractor Operators	27	25	1	0.7776
Butchers and Meat Cutters	30	37		0.7870
Payroll and Timekeeping Clerks	36	39		0.7924
Watch Repairers	36	39		0.7924
Hunters and Trappers	41	30	0	0.8063
Couriers and Messengers	32	37	1	0.8088
Forest and Conservation Workers	32	37		0.8088
Parking Lot Attendants	32	37	1	0.8088
Models	32	27		0.8117
Construction Laborers	29	30		0.8148
Locomotive Engineers	30	34		0.8148
Manicurists and Pedicurists	32	36		0.8202
Power Plant Operators	38	37		0.8267
Bakers	38	30		0.8325
Electrical and Electronic Equipment Assemblers	32	34	1	0.8351
Fishers and Related Fishing Workers	32	32		0.8402
Data Entry Keyers	34	30	1	0.8415
File Clerks	36	34	1	0.8553