

0.1 Introduction

The question of how computers can change future careers have been a popular issue recently. With the fast development of automation technologies, more jobs have become purely automated. Many people argued that the increase in unemployment rate is partly because this. Would it be true that more people will become jobless? Or is it just a short term phenomenon that does not represent future trends? This report represent how this problem is tackled in the respect of machine learning techniques. In particular, the property of jobs are analysed through Gaussian Process and the possibility of automation of the job can be predicted.

0.2 Gaussian Process Regression

0.2.1 Covariance Function

The covariance function describes how observations are related to each other.

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{(x - x')^2}{2l^2}\right]$$

0.2.2 Hyperparameters

0.3 Gaussian Process Classification

From the last chapter, we know that regression involves continuous observation values. However, sometimes the output data is not continuous but a series of labels, where we wish to determine the class labels according to input x . This is the problem of classification. Similar ideas are adopted, except here regression is applied on a latent variable f and then the final class label is determined by 'squashing' f into range $[0, 1]$, from which the probability of a typical label.

0.3.1 'Squashing' Function

The 'squashing' function can be any sigmoid function. Two typical sigmoid functions that will be discussed are logistic function $\lambda(f)$ and cumulative Gaussian function $\phi(f)$.

Inference is hence divided into two steps. First is computing the distribution of the prediction latent variable f_* in terms of previous observations

$$p(f_* | X, y, x_*) = \int p(f_* | X, x_*, f) p(f | X, y) df$$

where $p(f | X, y) = p(y | f) p(f | X) / p(y | X)$ is the posterior of latent variables f , which is estimated by MAP with respect to hyperparameters (μ and σ_f).

After obtaining the information about predictive latent variable, the probabilistic prediction and estimated by

$$\bar{\pi}_* = p(y_* = +1 | X, y, x_*) = \int \sigma(f_*) p(f_* | X, y, x_*) df_*$$

This is called the averaged probability. One may argue that expectation of the prediction should be simply equal to the sigmoidal mean of f_* . These are actually two different things because of the non-linearity of sigmoid function, the first is $E[\pi_* | X, y, x_*]$ while the later one is $\sigma(E[f_* | y])$.

0.3.2 The Laplace Approximation

Both of the integrals mentioned above are not analytically tractable. By doing a second order Taylor expansion of $\log p(f | X, y)$ around its maximum point, a Gaussian approximation of the posterior $q(f | X, y)$ can be obtained.

0.3.3 Estimating the Model

0.3.4 Probabilistic Prediction

0.3.5 Mathematical Implementation

Optimised latent variable f

The best latent f should be estimated for each set of hyperparameters by solving

$$\hat{f} = K(\nabla \log p(y|\hat{f}))$$

Commonly used convergence criteria depend on the difference between successive $\Psi(f)$, the magnitude of gradient vector $\nabla \Psi(f)$ or the difference between successive values of f . Note in practice, the convergence of objective function is assured by checking that each iteration gives an increase in $\Psi(f)$. If not, a smaller step change in f should be used.

Maximised Marginal Likelihood

Gradient computation:

Predictive Class Probability

Adding jitter -

0.3.6 Multi-class Classification