MATH97131– Machine Learning
Coursework 1 — Spring 2022

Submit by 6pm Friday the 4th of February 2022. Upload your final version as a PDF file with at most 10 pages (excluding the appendix). Avoid last minute uploads.

As this is assessed work you need to work on it individually. It must be your own and unaided work. You are not allowed to discuss the assessed coursework with your fellow students or anybody else. All rules regarding academic integrity and plagiarism apply. Violations of this will be treated as an examination offence. All questions that you may have concerning the coursework must be addressed to the lecturer via e-mail (marking the e-mail as high priority). Any resulting clarifications will be communicated to the entire year via Blackboard announcements.

Please note the following:

- Considerable emphasis will be put on clarity of expression, quality of presentation and on the depth of understanding. Ensure that your answers are well written, organised and are in the form of properly written sentences that include your full statistical reasoning. Use mathematical equations to describe your reasoning.
- Marks will be allocated to the quality of the presentation including the quality of the figures.
- Please clearly state in your essay the name of the `R` or `Python` packages and functions you are using. Provide your code in appendix – do not use any code in your essay.
- Report results rounded with 4 digits.

Question 1 – 50% of total mark

Please state the last two digits of your CID at the beginning of the report. Download your individual dataset $\mathcal{D} = \{x_i, z_i, y_i\}_{i=1}^{100}$ available on Blackboard. Consider the following model

$$Y_i = aZ_i + (b_1 + b_2 Z_i)\cos(X_i) + (c_1 + c_2 Z_i)X_i + (d_1 + d_2 Z_i)X_i^2 + (e_1 + e_2 Z_i)X_i^3 + \epsilon_i$$

where ϵ_i are independent random variables such that $\mathbb{E}(\epsilon_i) = 0$.

1. Infer the parameters of the model using ridge regression. Precisely describe the fitting procedures including the choice of the regularisation parameter. Report the inferred parameters. Use the inferred model to predict the value of Y for $X \in \{-7, -6.9, -6.8, \dots 6.8, 6.9, 7\}$ (which is an equally spaced grid between -7 and 7) and $Z \in \{0, 1\}$. Produce two plots comparing the predictions with the observed data $\mathcal{D}$; one plot for $z = 0$ and another one for $z = 1$. Comment on the fit of the model.
2. Suppose now that $\epsilon_i \sim \mathcal{N}(0, 0.1)$. Assuming that the parameters are *a priori* independent and considering a Gaussian prior distribution with mean 0 and variance α for each parameter, perform a Bayesian inference of the model parameters. Plot the posterior mean of the 9 model parameters as a function of α. Discuss the results. Compute the posterior predictive distribution for any values of $X \in \{-7, -6.9, -6.8, \dots 6.8, 6.9, 7\}$ and $Z \in \{0, 1\}$ for $\alpha = 1$. Compare the posterior prediction with the observed data $\mathcal{D}$ by producing two plots (one for $z = 0$ and one for $z = 1$) showing the mean of the posterior predictive distribution as a function of X as well as a 95% credible interval. Comment on the fit of the model.
3. Suppose that we are interested in making prediction for the following four observations $(X, Z) \in \{-5, 7\} \times \{0, 1\}$. Compare the posterior predictive distributions of the Bayesian model from question 2 using two appropriately chosen values of α as well as the predictions using the inferred ridge regression model from question 1.

Question 2 – 50% of total mark

Consider the Million Song Data Set described at `https://archive.ics.uci.edu/ml/datasets/YearPredictionMSD`. In this question you will use a subset of a slightly modified version of this dataset which is available on Blackboard. For each song, the dataset contains the year of the song as well as 90 attributes. The aim of this question is to classify whether a song has been produced before or after 1980 using three different classifiers:

- a LDA classifier
- a k-nearest neighbours classifier
- and a Naive Bayes classifier (with appropriately chosen distributions).

Write a small report describing the dataset, the proposed classifiers and comparing their performances (in terms of error rates as well as ROC curves and computational cost) for classifying the songs in this dataset. Precisely explain how the classifiers are constructed (using mathematical equations and/or pseudo-code) stating the modelling assumptions. Provide details regarding the training and testing procedure and describe the procedure to tune any hyper-parameters.