COMP3032 – Machne Learning Assignment One (20 marks)

Marking guide

Task1 (12 marks):

Polynomial regression models

1 Read data and create models (2)

Read datasets *pressure.csv* properly, and assign the correct columns to the respective variables properly. Create polynomial regression models

2 Plot (2)

Produce a cross validation error plot using the mean RMSE with 1 to 14 different degrees

3 Print coefficients(1)

For the best degree of your choice, print its coefficients.

Multiple linear regression model

4 Read data and create model (2)

Read datasets, and create a multiple linear regression model to predict systolic pressure using all the other relevant features. Print its coefficients

5 Cross validation (1)

Perform 10-fold cross validation. Display the mean RMSE

Ridge regression model

6 Ridge regression model (2)

Build the ridge regression model. Print its coefficients. Perform 10-fold cross validation and display the mean RMSE

Short answer and documentation

7 short answer and documentation (2)

Select the best degree, and explains why briefly. Select the best model of the three, and explains why briefly. Other Documentation (logic description, test run output, comments etc)

Task2 (8 marks):

1. Read data (1)

   Read dataset 'mnist_784' and assign the correct columns to the respective variables properly. Transform the label properly according to the requirement.

2. PCA (2)

Prform Principal Component Analysis (PCA) on the feature data to reduce its dimensionality while retaining 88% of the overall explained variance ratio. Display the number of principal components preserved.

3. Logistic model and Predictions (2)

   Create a logistic regression model using the reduced dataset. Dataset splits properly.

   Use this model to predict the language for the training set and testing set

4. Accuracy (2)

   Calculate the accuracy , confusion matrix, miss classified digits and display them.

5. Short answer and documentation (1)

   What do you think of the model generated (good, underfit, overfit)? Briefly explains why. Other Documentation (logic description, test run output, comments etc)