

Indian Institute of Technology Dharwad
Department of EE/CSE/ME
EE 405: Pattern Recognition and Machine Learning

Total Marks: 10

Quiz 1

Date: 06/10/2020

Instructions:

- Use meaningful variable names and write appropriate comments wherever required.
 - Use appropriate libraries to read, plot and manipulate the data.
 - Ensure that all the libraries are included in the final version of file that will be submitted.
 - Any genuine queries regarding any question must be posted in moodle. Queries to any of the TAs through WhatsApp or email will not be entertained.
-

- In this exercise your task is to implement a linear regression model from scratch on the given dataset. Obtain the parameters by both matrix inversion and gradient descent methods.
- The dataset has 3 features and 1 target variable. Before starting any implementation, it is often recommended to visualize and understand the data distribution. For the given dataset, scatter plot can be used as a visualization tool. As 2-D figures are easier to comprehend compared to 3-D, plot the target variable w.r.t to each feature to observe the trend.

- **Note:-** For each of the models below you have to perform k -fold cross validation (CV) and obtain the train and test split to build and validate the model.

To perform k -fold CV, randomize the data, split the data into approximately k equal parts, consider $k - 1$ parts for training and one for testing. Repeat the loop k times.

- For example, to perform 5-fold CV:

You have divided your dataset into 5 equal parts (1, 2, 3, 4 and 5). Take four parts at a time for training and one for testing.

- Dataset 1: Training set- 1, 2, 3, 4 and Testing set- 5.
- Dataset 2: Training set- 1, 2, 3, 5 and Testing set- 4.
- Dataset 3: Training set- 1, 2, 4, 5 and Testing set- 3.
- Dataset 4: Training set- 1, 2, 3, 5 and Testing set- 2.
- Dataset 5: Training set- 2, 3, 4, 5 and Testing set- 1.

1. **(4 Marks)** Observe the data by plotting target variable w.r.t each feature and based on your observation comment whether you will use linear regression or polynomial regression. If you choose polynomial regression then what is its degree? Now, validate your observation computationally by performing the task given below.

Perform univariate linear regression task w.r.t each feature and tabulate the following:

- (a) Estimated parameters and errors for each fold.
- (b) Comment on the variation obtained in the values of estimated parameters and error in each fold.

Similarly, perform polynomial regression and tabulate the results. Compare linear and polynomial regression and comment on the order of polynomial regression which perfectly fits the data. Plot the best fit line w.r.t each feature.

2. **(4 Marks)** Observe the data by plotting target variable by taking 2 features at a time and based on your observation comment whether you will use linear regression or polynomial regression. If you choose polynomial regression then what is its degree? Now, validate your observation computationally by performing the task given below.

Perform bivariate regression task by taking 2 features at a time and tabulate the following:

- (a) Estimated parameters and errors for each fold.
- (b) Comment on the variation obtained in the values of estimated parameters and error in each fold.

Similarly, perform polynomial bivariate regression and tabulate the results. Compare linear and polynomial regression and comment on the order of polynomial regression which perfectly fit the data.

Plot the best surface w.r.t each feature vector.

3. **(2 Marks)** Perform regression analysis considering all the 3 features and tabulate the following:

- (a) Estimated parameters and errors for each fold.
- (b) Comment on the variation obtained in the values of estimated parameters and error in each fold.