


ME 453: Data Science in Manufacturing Quality Control

Homework 6

Assigned: December 8, 2023

Due: December 18, 2023


Homework guidelines:

1. The total number of points is 100. There are also 10 bonus points available from the bonus problems. The assigned points are given beside questions. To get full credit you must **SHOW ALL OF YOUR WORK**.
2. Bonus points will be used to compensate for lost points in homework assignments.
3.  Indicates the problem which you need to use Python.
4. A complete submission features the following items: (a) a brief report including all figures and results, and explanations of necessary steps taken to obtain them; and (b) and the source code (Python is recommended).
5. Item (a) can be scanned copies of handwritten or typeset reports. Both items shall be submitted through Canvas.

Review the following topics discussed in lecture:

- Topic 11 - Classification

Problem 1 (100 points + 10 points)

 This problem is a continuation of HW5. We use the same bearing data set to study the performance of different classifiers. There are four wear levels referred to as level0-3. Each level has 30 samples. 5 features have been extracted from the raw signals. All features have been normalized and anonymized. The data has been randomly partitioned into a training set and a test set, which are available in the csv files.

- (1) Calculate Fisher's ratio for "level 0 - level 3" (Fisher's ratio 1), "level 1 - level 2" (Fisher's ratio 2), and compound Fisher's ratio (Fisher's ratio 1 + Fisher's ratio 2) for all features. (10 points)
- (2) Plot the compound Fisher's ratio vs. feature index in a bar plot. (5 points)
- (3) Build an LDA and a QDA classifier using all features. Report the confusion matrix and accuracy for testing data set. (20 points)
- (4) Build an LDA and a QDA classifier using two features with the highest compound Fisher's ratios. Report the confusion matrix and accuracy for the testing data set. (20 points)

- (5) For LDA and QDA, does adding more feature improve the classification performance? (5 points)
- (6) Build an SVM classifier using two features with the highest compound Fisher's ratio. Experiment with the following kernels: (i) 1st degree polynomials (`linear` in Python), (ii) 2nd degree polynomials (`poly`), and (iii) radial basis function (`rbf`). For each kernel, report the confusion matrix and accuracy for testing data set. (20 points)
- (7) Build KNN classifier with $K = 1$ and $K = 5$. For each K , report the confusion matrix and accuracy for testing data set. (20 points)
- (8) [Bonus] Compare the results from (4), (7), and (8). Which classifier is better for this data set? Justify your answer. (10 points)