ME 453: Data Science in Manufacturing Quality Control
Homework 5

Assigned: November 24, 2023 Due: December 8, 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:

- Quality Control in Today's Manufacturing (lecture videos 17)
- Feature Generation (lecture videos 18)

Problem 1 (20 points + 4 points)

Suppose the estimated covariance matrix for a data set is:

$$\hat{\Sigma} = \begin{bmatrix} 16 & 5 \\ 5 & 9 \end{bmatrix}$$

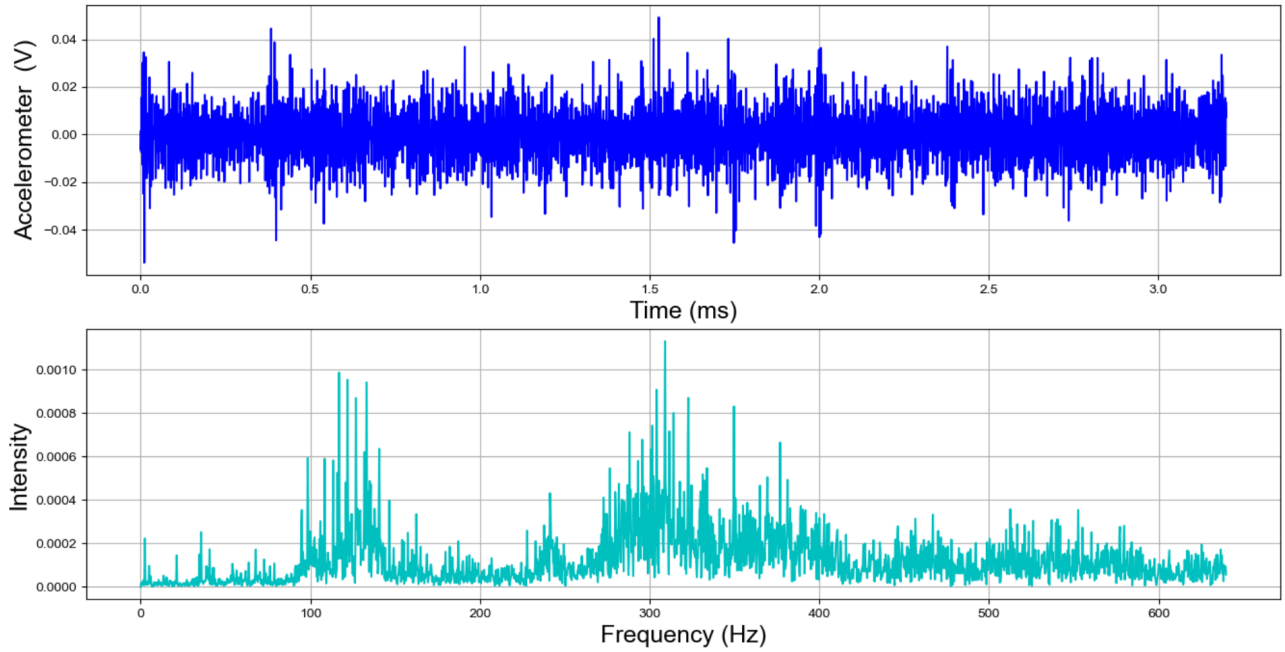
(1) Obtain the projection matrix W, which is no other than eigenvectors stacked horizontally. (15 points)

(2) Project an observation $x = [5 \quad 4]^T$ to the newly mapped space. (5 points)

(3) [Bonus] Calculate the explained variance ratio for each PC (principal component). Which PC is more important? Justify your answer. (4 points)

Problem 2 (80 points + 6 points)

In the `bearing.csv` provided on Canvas, the accelerometer signal is collected during a rotational test of bearing. A raw signal is plotted with FFT result in the figure below. We are interested in the effect of bearing ball wear, which has four level: slight, light, moderate, and severe. The wearing levels are stored in `wearlevel.csv` and referred as level 0-3 correspondingly.



(1) Perform PCA on the dataset. Note that the mean should be removed at each time step (centered) for best results.

   (a) Plot scree plot of 30 PCs and determine the number of components to keep. Use $\alpha = 0.1$. (15 points)

   (b) Visualize the data with the first three PCs in a 3D scatter plot. Distinguish the wear levels by different colors and provide a legend. (10 points)

(2) Perform FFT and extract two frequency-domain features and provide a description for each feature. (15 points)

(3) Visualize the data using handcrafted features obtained in (2) in a scatter plot. Distinguish the wear levels by different colors and provide a legend. (10 points)

(4) Create a feature pool includes two frequency-domain features and first three PCs. Calculate Fisher's ratio for "level 0 - level 3" (Fisher's ratio 1) and "level 1 - level 2" (Fisher's ratio 2) for all features. (20 points)

(5) Select three features from the feature pool based on the following criterion: max (Fisher's ratio 1 + Fisher's ratio 2). Visualize the three selected features in a 3D plot. Use legend to distinguish the data points from different wear levels. (10 points)

(6) [Bonus] Use the intensity from FFT as the data, remove the mean (centered) and perform PCA. Plot scree plot of 30 PCs and determine the number of components to keep. Use $\alpha = 0.1$. (3 points)

(7) [Bonus] Compare and discuss the PCA results from (1) and (6) when using raw signal. (3 points)