For Questions 1- 2, please submit a word file or a PDF file;
For Question 3 (programming question), please submit an .ipynb file.

Question 1 (25 points):

1) Explain why it is important to reduce the dimension and remove irrelevant features of data (e.g., using PCA) for Instance-Based Learning such as kNN? (5 points)
2) One limitation with K-Means is the variability issue. Explain how to address this problem. (5 points)
3) Please explain the technique of Gaussian Mixture and how it is used for anomaly detection. (5 points)
4) Please draw the diagram of Convolutional Neural Networks (CNN). Then explain the functionality of each layer of CNN. Name several latest algorithms of CNN (e.g., AlexNet etc.). (5 points)
5) What are the vanishing and exploding gradients problems in Backpropagation? Name several techniques to address these problems. (5 points)

Question 2 (5 points):

Consider a learned hypothesis, *h*, for some Boolean concept. When *h* is tested on a set of 100 examples, it classifies 80 correctly. What is the 95% confidence interval for the true error rate for $Error_D(h)$?

Question 3 – Programming (20 points):

Design a genetic algorithm to solve the polynomial fitting problem that we did in Homework #1. You need to implement a genetic algorithm using BOTH **mutation** AND **crossover** operations. You need to decide a mutation rate and a crossover rate.

Plot the following in one figure: 1) the original noisy data, 2) the polynomial you obtained in Homework #1, and 3) the polynomial obtained from this implementation. Compare and discussion the difference in performance of the two polynomials obtained with two different methods.