QUIZ 1: 7:00-7:40 pm (*60 points total*)

Note: This is a **closed-book** exam. Please write your answers on clean sheets of letter paper using a pen or pencil, and upload the scanned pages on Gradescope, indicating the page numbers corresponding to your answers when you upload on Gradescope.

1. [**Bayes Classifier (10 points)**]
   Consider the binary hypothesis testing problem:

$$H_{-1}: \ X \sim p(x| -1)$$
$$H_1: \ X \sim p(x|1)$$

where

$$p(x|-1) = \begin{cases} 0.1 & x = 0 \\ 0.3 & x = 1 \\ 0.5 & x = 2 \\ 0.1 & x = 3 \\ 0 & \text{otherwise} \end{cases} \qquad p(x|1) = \begin{cases} 0.6 & x = 0 \\ 0.1 & x = 1 \\ 0.1 & x = 2 \\ 0.2 & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

with priors $\pi_{-1} = 0.2, \pi_1 = 0.8$.

   (a) Find the Bayes classifier (MAP decision rule).

   (b) Calculate the probability of error for the Bayes classifier.

2. [**Linear Discriminant Functions(10 points)**]
   Consider a 3-ary linear classifier, with classes 1, 2, and 3, for which the three linear discriminant functions are:

$$g_{12}(\underline{x}) = x_1 + x_2 - 2$$
$$g_{13}(\underline{x}) = 2x_1 - x_2 - 3$$
$$g_{23}(\underline{x}) = x_1 - 2x_2 - 1$$

   Classify the input $\underline{x}$ with $x_1 = 1, x_2 = 5$.

3. [**Logistic Regression (10 points)**]
   In (binary) logistic regression we model the probabilities of the labels given the feature vector as:

$$p(1|\underline{x}) = \frac{\exp(\beta_0 + \underline{\beta}^\top \underline{x})}{1 + \exp(\beta_0 + \underline{\beta}^\top \underline{x})}, \quad p(-1|\underline{x}) = \frac{1}{1 + \exp(\beta_0 + \underline{\beta}^\top \underline{x})}$$

   where β_0 and $\underline{\beta}$ are estimated from the training data.

   Show that such a model results in a linear classifier.

4. [**Concepts (30 points)**]
   Answer in one or two complete sentences.

   (a) What is the difference between prediction error and validation error?

   (b) The Bayes classifier minimizes the prediction error. Why can we not use it in supervised learning?

   (c) What underlying assumption is made about the distribution of the feature vectors under each class in constructing the LDA classifier?

   (d) Explain, using a figure if necessary, how a binary SVM classifier is designed for linearly separable data.

   (e) Write down *one* of the advantages (pros) of the Naive Bayes classifier.

   (f) What is the difference between Model Selection and Model Assessment?