ECE 411 Introduction to Machine Learning
Fall 2024 Exam 2
Instructor: Dr. Chau-Wai Wong

This is a closed-book exam. You may use a scientific calculator with cleared memory, but not a smart phone or computer. Two-sided letter-sized handwritten cheatsheet is allowed. You should answer ***all four*** problems.

Problem 1 [ML Concepts] (26 pts) True or False. If a sentence is true, write "True" on the answer sheet (2 pts). Otherwise, write "False" (1 pt) and copy a few keywords (at most 3 words) responsible for making it a false statement (1 pt).

1. When viewed as generalized linear regression, logistic regression inherits the pros and cons of linear regression.
2. Maximum likelihood estimation leverages the degree of match between the training data and the tuneable model to estimate the model parameters led to the generation of the training data.
3. Discriminant analysis has to assume a Gaussian distribution for data points of each class.
4. Discriminant analysis involving multivariate Gaussian distributions can have planar/flat decision boundaries if the Gaussian parameters fulfill specific constraints.
5. For imbalanced binary classification problems, an equal error rate (EER) or area under the curve (AUC) cannot be a meaningful indicator of the system's classification performance because either of them is merely a single number.
6. Discriminant analysis and logistic regression are discriminant learning methods, whereas generative pretrained transformers (GPTs) are generative learning methods.
7. k-fold cross-validation is an estimator for the true prediction error. This estimator has a larger variance than the leave-one-out cross-validation (LOOCV) because it uses fewer folds than LOOCV.
8. A bootstrap sample $\{5, 2, 2, 0, -1\}$ could be constructed by drawing 5 data points one by one from a random sample $\{5, 2, -1, 3, 0\}$ with replacement.
9. The chance of obtaining bootstrap sample $\{2, 2, 2, 5, 5\}$ from a random sample $\{5, 2, -1, 3, 0\}$ is $\binom{5}{2}(1/5)^5$, where $\binom{5}{2}$ corresponds to the number of scenarios that two distinct 5s are inserted in two of the five candidate locations.

10. As the value of a nonnegative regularization parameter increases, every ridge regression coefficient estimate becomes smaller in magnitude.

11. As the value of a nonnegative regularization parameter increases, the flexibility of the ridge regression model decreases.

12. Both lasso and ridge can make some estimated coefficients exactly zero.

13. For any two-class training dataset, a support vector machine can narrow down to a few support vectors to create a decision boundary with two margin lines around it.

Problem 2 (24 pts) This problem concerns the maximum likelihood estimator (MLE). Use the invariance principle of MLE if needed.

(a) Calculate the MLE for variance σ^2 for a random sample $X_1, \ldots, X_n$ drawn from the normal distribution with the PDF shown as follows:

$$f(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty. \tag{1}$$

(b) Calculate the MLE for parameter θ for a random sample $X_1, \ldots, X_n$ drawn from an exponential distribution with the PDF of the following form:

$$f(x; \theta) = \frac{1}{\theta} \, e^{-x/\theta}, \quad x \geq 0. \tag{2}$$

Problem 3 (25 pts) This problem concerns the curse of dimensionality.

(a) Suppose that we have a set of observations, each with measurements on $p = 1$ feature, X. We assume that X is uniformly distributed on $[-1, 1]$. Associated with each observation is a response value. Suppose that we wish to predict a test observation's response using only observations that are within 10% of the range of X closest to that test observation. For instance, in order to predict the response for a test observation with $X = 0.2$, we will use observations in the range $[0.1, 0.3]$. On average, what fraction of the available observations will we use to make the prediction for a test observation with $X = 0.3$?

(b) Now suppose that we have a set of observations, each with measurements on $p = 2$ features, X_1 and X_2. We assume that (X_1, X_2) are uniformly distributed on $[-1, 1] \times [-1, 1]$. We wish to predict a test observation's response using only observations that are within 20% of the range of X_1 and within 20% of the range of X_2 closest to that test observation. On average, what fraction of the available observations will we use to make the prediction for a test observation with $(X_1, X_2) = (0.1\pi, 0.2\pi)$?

(c) Generalize the cases in (a) and (b) to $p = 100$. What fraction of the available observations will we use to make the prediction? We assume that the test observation stays far away from the boundaries.

(d) Using your answers to (a)–(c), comment on a drawback of k-NN when p is large.

(e) Now suppose that we wish to make a prediction for a test observation by creating a p-dimensional hypercube centered around the test observation that contains, on average, 20% of the training observations. For $p = 1, 2,$ and 100, what is the length of each side of the hypercube?

Problem 4 (25 pts) This problem concerns the application of the method of nearest neighbors. A set of 6 training data points is drawn in the given figure. Using the k-nearest-neighbor regression rule, an estimated regression function can be written as follows:

$$\hat{y}^{(k)} = \hat{f}^{(k)}(x) = \frac{1}{k} \sum_{i \in I(x)} y_i, \quad I(x) = \{i : \text{the indices of } k \text{ smallest } |x_i - x|\}. \tag{3}$$

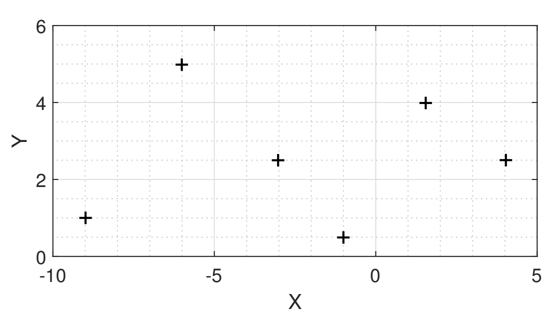
Note that this is a regression problem with only one predictor and the graphical representation of the estimated regression function on the xy-plane will be a collection of horizontal line segments.

(a) Draw the regression function $\hat{f}^{(k)}(x)$ for $x \in [-10, 5]$ when only $k = 1$ nearest neighbor is contributing to the regression.

(b) Draw the regression function $\hat{f}^{(k)}(x)$ for $x \in [-10, 5]$ when two $k = 2$ nearest neighbors are contributing to the regression.

(c) Comment on how the shape of regression function will change as the number of contributing neighbors increases.

To get full points, you must annotate the locations of the ***discontinuities*** of each estimated regression function using vertical dotted lines.



[This Page Intentionally Left Blank]

Answer sheet for Problem 1 Name: Obtained Points:

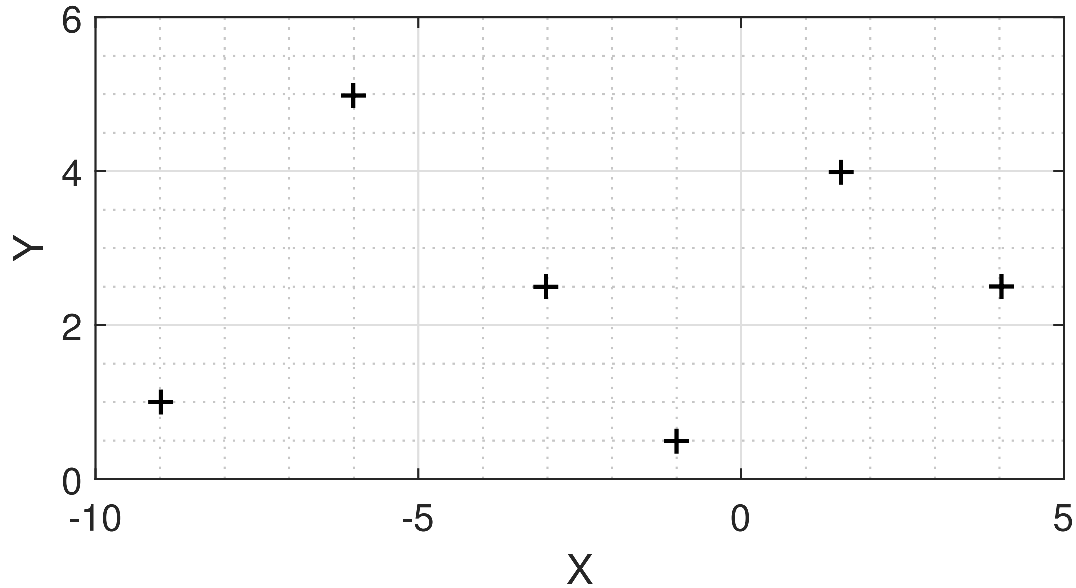
Answer sheet for Problem 1

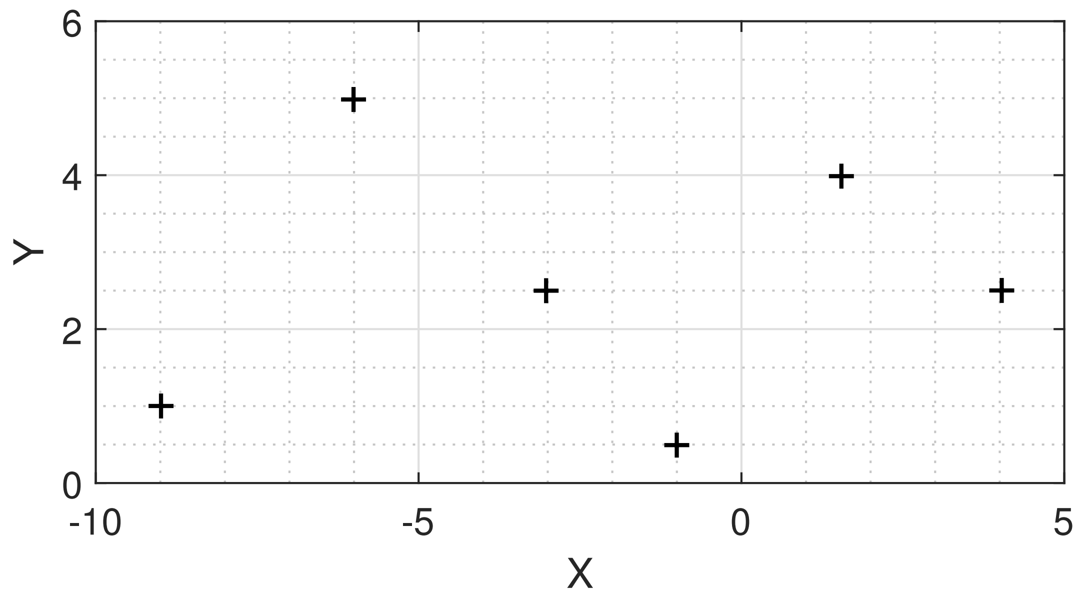
Answer sheet for Problem 2 Name: Obtained Points:

Answer sheet for Problem 2

Answer sheet for Problem 3 Name: Obtained Points:

Answer sheet for Problem 3


6
4
2
0
Y
-10
-5
0
5
X


6
4
2
0
Y
-10
-5
0
5
X

Answer sheet for Problem 4