

# CSCI 416/516 Practice Midterm

Name:

Student ID:

---

**Submission:** Write your name and student id. You have 1 hour and 20 minutes to complete your exam. You are allowed a one-sided cheat sheet and a calculator.

- **Problem 1 [2pts]: Algorithm Category.**

Which of the following algorithms is non-parametric?

- Logistic Regression
- Linear Regression
- K-Nearest Neighbors (KNN)
- None of the above

- **Problem 2 [2pts]: Logistic Regression.**

In logistic regression, what is the range of values that the hypothesis can take?

- 0 to 1
- $-\infty$  to  $\infty$
- 0 to 100
- -1 to 1

- **Problem 3 [2pts]: Linear Regression.**

The primary purpose of linear regression is to:

- Classify data into two or more categories
- Predict a continuous-valued output
- Cluster data into distinct groups
- Reduce the dimensionality of data

- **Problem 4 [2pts]: KNN.**

In the context of KNN, what does K represent?

- The number of features
- The number of data points
- The number of nearest neighbors

– The classification label

- **Problem 5 [2pts]: Linear Regression.**

What is the loss function used in linear regression? Describe its significance.

- **Problem 6 [2pts]: Euclidean Distance.**

Consider the following 3-dimensional points,  $x^{(a)} = [5, 9, -3]$  and  $x^{(b)} = [1, 2, -6]$ . Write the formula for the Euclidean distance between two points in a 3-dimensional space. Then, using the formula, calculate the Euclidean distance between  $x^{(a)}$  and  $x^{(b)}$ .

- **Problem 7 [3pts]: Linear Regression.** Given the linear regression equation  $y = w_0 + w_1x_1 + w_2x_2$ , describe what each  $w$  represents.

- **Problem 8 [bonus 3pts]: Linear Regression.** Using the gradient descent algorithm, write out the update rules for the coefficients in a linear regression model.