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. **For all questions, multiple choice questions excluded, you need to show the calculations and/or steps that you take to reach the conclusion.**

- **Problem 1 [3 pts]: (Partial) Derivatives.**
  - (a) [1 pt] Solve the following partial derivative in the context of logistic regression. Rewrite the result into an expression solely with y.

$$\frac{\partial}{\partial z}\frac{1}{1+e^{(-z)}} \tag{1}$$

  - (b) [1 pt] Solve the following partial derivative in the context of linear regression, whereas $\boldsymbol{w} = \{w_1, w_2, ..., w_j, ..., w_d\}$ since $\boldsymbol{w}$ has a dimensionality of d.

$$\frac{\partial}{\partial w_j}\boldsymbol{w}^\top \boldsymbol{x} + b \tag{2}$$

  - (c) [1 pt] Solve the following partial derivative in the context of linear regression, whereas $\boldsymbol{w} = \{w_1, w_2, ..., w_j, ..., w_d\}$ since $\boldsymbol{w}$ has a dimensionality of d.

$$\frac{\partial}{\partial b}\boldsymbol{w}^\top \boldsymbol{x} + b \tag{3}$$

- **Problem 2 [3 pts]: KNN.**
  KNN typically uses Euclidean distance as its default metric.
  - (a) [2 pt] However, which of the following metrics CANNOT be used as the distance metric in KNN? (Hint: we covered this question in class.)
    - A. Kullback–Leibler (KL) Divergence
    - B. Cosine Similarity
    - C. Edit Distances (such as Hamming Distance)
    - D. All of the above can be used as the distance metric
  - (b) [1 pt] Why? Please explain your choice above.

- **Problem 3 [2 pt]: Gradient Descent.**
  Write down 1 critical flaw for the optimization algorithm gradient descent (including batch, stochastic, and mini-batch gradient descent).

- **Problem 4 [2 pt]: 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 [2 pt]: Linear Regression.**
  Using the gradient descent algorithm, write out the update rules for the parameters b and w_j in a linear regression model, given the learning rates α_{w_j} and α_b, and the regularized cost function $\boldsymbol{J}_{\text{reg}}$. You don't need to expand the expressions.

  Your answer should have the form:

$$w_j \leftarrow ... \tag{4}$$

$$b \leftarrow ... \tag{5}$$

- **Problem 6 [2 pts]: 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 [2 pts]: Linear Regression.** Given the linear regression equation $y = w_0 \times 1 + w_1 x_1$, describe what each w_n represents whereas $n = 0, 1$.

- **Problem 8 [4 pts]: Lagrangian Multiplier.**

  – (a) [1 pt] Suppose a sample in the training dataset has a Lagrangian multiplier of 0. What does this say about this sample?
  – (b) [1 pt] Suppose a sample in the training dataset has a Lagrangian multiplier of 1. What does this say about this sample?
  – (c) [1 pt] Can we have Lagrangian Multiplier being negative? Yes or No?
  – (d) [1 pt] Justify your answer to 5(c).

- **Problem 9 (Bonus) [3 pt]: Linear Regression Using Gradient Descent.**
  Given $\boldsymbol{x} = \{x_1, x_2, x_3, x_4\} = \{1, 2, 3, 4\}$ and $\boldsymbol{t} = \{t_1, t_2, t_3, t_4\} = \{10, 20, 30, 40\}$, and the initial $\boldsymbol{w}_{iter=0} = \{w_0, w_1\} = \{0.1, 0.1\}$ in which the bias b is incorporated as w_0, what is the weight $\boldsymbol{w}_{iter=1}$ after 1 iteration with a learning rate of 0.1?