CS-233 Theoretical Exercise

1. Logistic regression can be used to

a) predict tomorrow's temperature from previous daily temperature observation.

b) predict whether the object is glass or stone given its reflectance.

c) predict someone's weight based on their height.

d) predict the probability of raining given the humidity, wind force and temperature.

e) tell apart a dog from a cat given height and weight measurements.

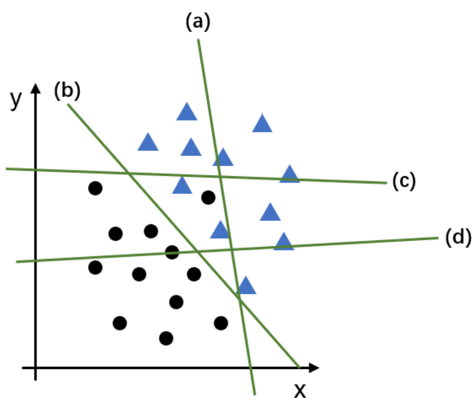
2. Which of the following statements are true:

a) Linear regression cannot be used for classification problems.

b) Logistic regression is a linear classifier and can only separate classes using linear decision boundaries.

c) The output of logistic regression is the estimated probability of the sample belonging to a specific class.

3. In the following figure, the dots and triangles are samples from two different classes. Which line is the most likely decision boundary obtained by logistic regression?



4. Consider a dataset with the four data points shown in Table 1. Assume $x^{(1)}$ and $x^{(2)}$ are two measured biochemical indicators of patients, and $y = 0$ and $y = 1$ indicate the patients without and with a specific symptom, respectively. We want to build a logistic regression model to predict whether a patient has the symptom based on the input features $x^{(1)}$ and $x^{(2)}$. The prediction model is expressed as

$$\hat{y} = \sigma(\mathbf{w}^T\mathbf{x}), \tag{1}$$

where σ is the sigmoid function, $\mathbf{w} = [w^{(0)}, w^{(1)}, w^{(2)}]^T$, $\mathbf{x} = [1, x^{(1)}, x^{(2)}]^T$.

(1) Write down the algorithm (pseudo-code) that uses gradient descent to compute the optimal $\mathbf{w}$.

(2) Perform one iteration of the previous algorithm with an initialization of $\mathbf{w} = [1, 1, 1]^T$ and a step size of 0.1.

(3) Assume $\mathbf{w}^*$ is the optimal solution obtained after the full gradient descent algorithm (not just a single step) in Question (2). Now, we divide $x^{(1)}$ and $x^{(2)}$ by 100 and perform the gradient descent algorithm

Data Point	$x^{(1)}$	$x^{(2)}$	y
1	-10	-20	0
2	0	-10	0
3	10	0	1
4	20	10	1

Table 1: Data points.

again with the scaled data. This results in a model with different parameters $\mathbf{w}'$. Given the test data of new patients $\{\mathbf{x}_5, ..., \mathbf{x}_N\}$, will the two classifiers defined by $\mathbf{w}^*$ and $\mathbf{w}'$ produce different results for them? Justify your answer mathematically. Note that you need to also scale the test data when using the classifier defined by $\mathbf{w}'$.

(4) If we switch the meaning of the y value, i.e., 0 and 1 now indicating with and without the symptom, respectively, and train a logistic regression model on the resulting data, will the model produce different results for $\{\mathbf{x}_5, ..., \mathbf{x}_N\}$ than before switching?