Question 1
Which function does logistic regression use to “squeeze” the real line to [0,1]?

a. Logistic function
b. Zero function
c. Absolute value function

Question 2
Consider training a 1 vs. all multi-class classifier for the problem of digit recognition using logistic regression. There are 10 digits, thus there are 10 classes. How many logistic regression classifiers will we have to train?

a. 9
b. 5
c. 10
d. 1

Question 3
In order to train a logistic regression model, we find the weights that maximize the likelihood of the model.

a. True
b. False

Question 4
In the context of L2 regularized logistic regression, which of the following occurs as we increase the L2 penalty λ? Choose all that apply.

a. Decision boundary becomes less complex
b. Training error decreases
c. The L2 norm of the set of coefficients gets smaller
d. Some features are excluded from the classifier
e. The classifier has lower variance
f. Region of uncertainty becomes narrower, i.e., the classifier makes predictions with higher confidence

Question 5
What is a multi-class classifier?

a. A classifier that can predict a field with two discrete values, such as ”Defaulter" or "Not Defaulter”.
b. A classifier that can predict a field with multiple discrete values, such as ”DrugA", ”DrugX" or "DrugY”.
c. A classifier that can predict multiple fields with many discrete values.

Question 6
You are training a classification model with logistic regression. Which of the following statements are true? Check all that apply.

a. Introducing regularization to the model always results in equal or better performance on the training set.
b. Adding a new feature to the model always results in equal or better performance on examples not in the training set.
c. Adding many new features to the model makes it more likely to overfit the training set.
d. Introducing regularization to the model always results in equal or better performance on examples not in the training set.

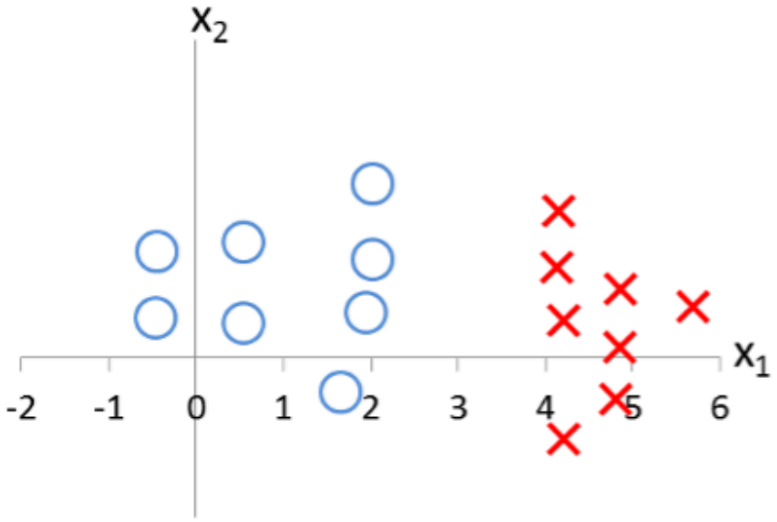
Question 7
Suppose you have implemented regularized logistic regression to classify what object is in an image (i.e., to do object recognition). However, when you test your hypothesis on a new set of images, you find that it makes unacceptably large errors with its predictions on the new images. However, your hypothesis performs well (has low error) on the training set. Which of the following are promising steps to take? Check all that apply.

a. Try using a smaller set of features
b. Use fewer training examples
c. Get more training examples
d. Try adding polynomial features

Question 8
Suppose you train an SVM and find it overfits your training data. Which of these would be a reasonable next step?

a. Decrease C
b. Increase C

Question 9
Consider the training set to the right, where "x" denotes positive examples (y=1) and "o" denotes negative examples (y=0). Suppose you train an SVM (which will predict 1 when θ0+θ1 x1+θ2 x2>= 0). What values might the SVM give for θ0, θ1, and θ2?



a. θ0 = 3, θ1 = 0, θ2 = 1
b. θ0 = -3, θ1 = 1, θ2 = 0
c. θ0 = 3, θ1 = 1, θ2 = 0
d. θ0 = -3, θ1 = 0, θ2 = 1

Question 10
Which of the following loss function is used by linear SVM?

a. Zero-one loss
b. Log loss
c. Hinge loss