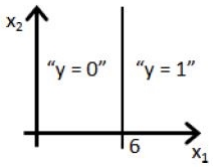
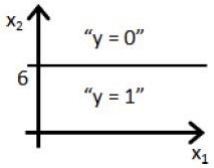
HMC CS 158
Quiz 4: Logistic Regression, Regularization

1. Suppose you train a logistic classifier $h_{\boldsymbol{\theta}}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$. Suppose $\theta_0 = 6$, $\theta_1 = 0$, $\theta_2 = -1$. Which of the following figures represents the decision boundary found by your classifier?

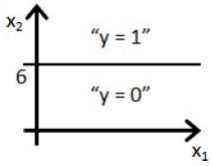
(a)



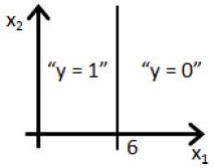
(c)



(b)

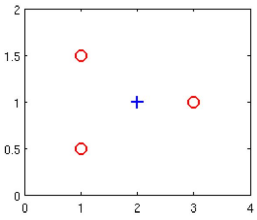


(d)



2. Suppose you have the following training set and fit a logistic regression classifier $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$.

x_1	x_2	y
1	0.5	0
1	1.5	0
2	1	1
3	1	0



Which of the following are true? Check all that apply.

(a) $J(\boldsymbol{\theta})$ will be a convex function, so gradient descent should converge to the global minimum.

(b) Adding polynomial features (e.g. instead using $h_{\boldsymbol{\theta}}(\boldsymbol{x}) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1 x_2 + \theta_5 x_2^2)$) could increase how well we can fit the training data.

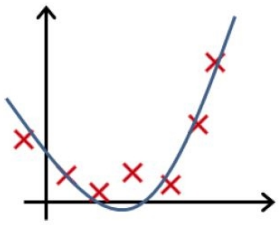
(c) Adding polynomial features (e.g. instead using $h_{\boldsymbol{\theta}}(\boldsymbol{x}) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1 x_2 + \theta_5 x_2^2)$) would increase $J(\boldsymbol{\theta})$ because we are now summing over more terms.

(d) Because the positive and negative examples cannot be separated using a straight line, linear regression will perform as well as logistic regression on this data.

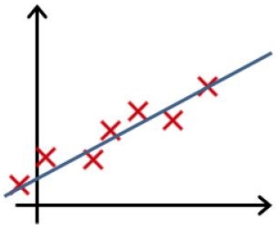
This quiz is adapted from course material by Andrew Ng (Stanford).

3. In which one of the following figures do you think the hypothesis has overfit the training set?

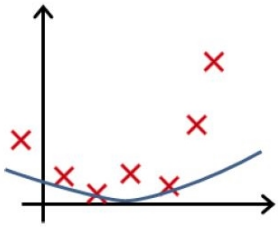
(a)



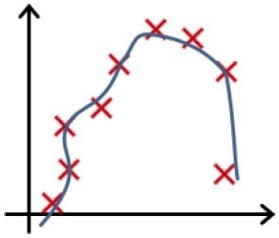
(c)



(b)

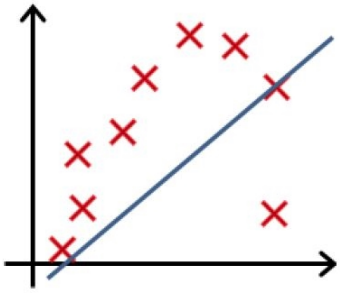


(d)

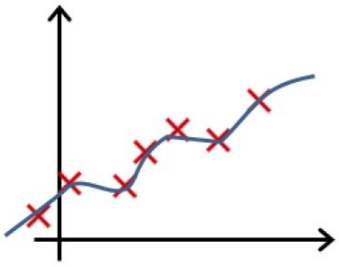


4. In which one of the following figures do you think the hypothesis has underfit the training set?

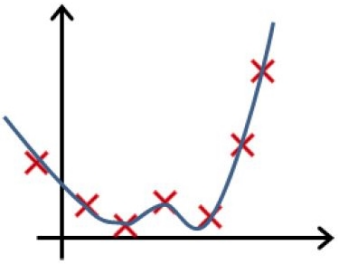
(a)



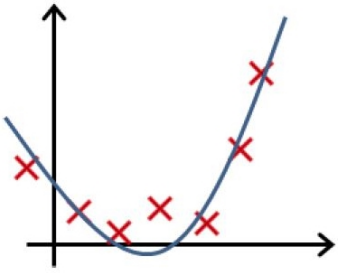
(c)



(b)



(d)



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

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

   (d) Introducing regularization to the model always results in equal or better performance on the training set.

6. Suppose you ran logistic regression twice, once with $\lambda = 0$ and once with $\lambda = 1$. One of the times, you got parameters $\boldsymbol{\theta} = \begin{pmatrix} 81.47 \\ 12.69 \end{pmatrix}$, and the other time you got $\boldsymbol{\theta} = \begin{pmatrix} 13.01 \\ 0.91 \end{pmatrix}$. However, you forgot which value of λ corresponds to which value of $\boldsymbol{\theta}$. Which one do you think corresponds to $\lambda = 1$?

   (a) $\boldsymbol{\theta} = \begin{pmatrix} 81.47 \\ 12.69 \end{pmatrix}$

   (b) $\boldsymbol{\theta} = \begin{pmatrix} 13.01 \\ 0.91 \end{pmatrix}$