Feedback — VI. Logistic Regression

Help

You submitted this quiz on **Mon 11 Nov 2013 2:09 PM PST**. You got a score of **4.50** out of **5.00**. You can attempt again in 10 minutes.

Question 1

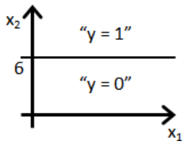
Suppose that you have trained a logistic regression classifier, and it outputs on a new example x a prediction $h_\theta(x)$ = 0.7. This means (check all that apply):

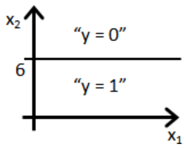
Your Answer		Score	Explanation
☐ Our estimate for $P(y=0\|x;\theta)$ is 0.7.	✔	0.25	$h_\theta(x)$ is $P(y=1\|x;\theta)$, not $P(y=0\|x;\theta)$.
☐ Our estimate for $P(y=1\|x;\theta)$ is 0.3.	✔	0.25	$h_\theta(x)$ gives $P(y=1\|x;\theta)$, not $1-P(y=1\|x;\theta)$.
☑ Our estimate for $P(y=1\|x;\theta)$ is 0.7.	✔	0.25	$h_\theta(x)$ is precisely $P(y=1\|x;\theta)$, so each is 0.7.
☑ Our estimate for $P(y=0\|x;\theta)$ is 0.3.	✔	0.25	Since we must have $P(y=0\|x;\theta)=1-P(y=1\|x;\theta)$, the former is 1 - 0.7 = 0.3.
Total		1.00 / 1.00	

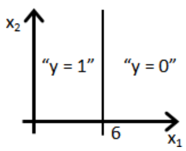
Question 2

Suppose you train a logistic classifier $h_\theta(x)=g(\theta_0+\theta_1x_1+\theta_2x_2)$. Suppose $\theta_0=6, \theta_1=-1, \theta_2=0$. Which of the following figures represents the decision boundary found by your classifier?

Your Answer	Score	Explanation

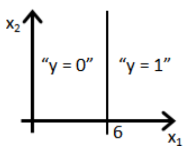






1.00

In this figure, we transition from negative to positive when x_1 goes from above 6 to below 6 which is true for the given values of θ.

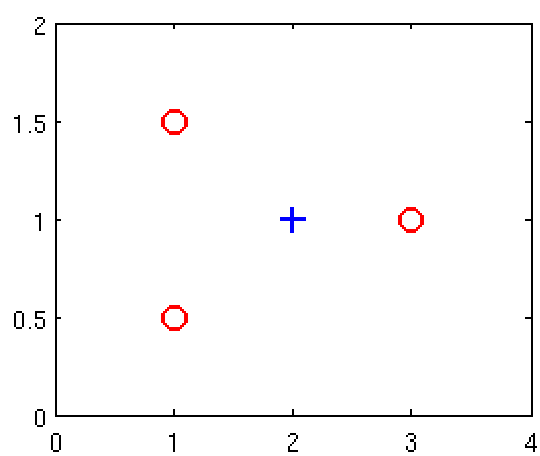


Total 1.00 / 1.00

Question 3

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.

Your Answer		Score	Explanation
☑ Adding polynomial features (e.g., instead using $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1 x_2 + \theta_5 x_2^2)$) could increase how well we can fit the training data.	✔	0.25	Adding new features can only improve the fit on the training set: since setting $\theta_3 = \theta_4 = \theta_5 = 0$ makes the hypothesis the same as the original one, gradient descent will use those features (by making the corresponding θ_j non-zero) only if doing so improves the training set fit.
☑ At the optimal value of θ (e.g., found by fminunc), we will have $J(\theta) \geq 0$.	✔	0.25	The cost function $J(\theta)$ is always non-negative for logistic regression.
☐ 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.	✔	0.25	While it is true they cannot be separated, logistic regression will outperform linear regression since its cost function focuses on classification, not prediction.
☐ The positive and negative examples cannot be separated using a straight line. So, gradient descent will fail to converge.	✔	0.25	While it is true they cannot be separated, gradient descent will still converge to the optimal fit. Some examples will remain misclassified at the optimum.
Total		1.00 /	

1.00

Question 4

For logistic regression, the gradient is given by $\frac{\partial}{\partial \theta_j} J(\theta) = \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$. Which of these is a correct gradient descent update for logistic regression with a learning rate of α? Check all that apply.

Your Answer		Score	Explanation
☑ $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)} \right) x_j^{(i)}$ (simultaneously update for all j).	✔	0.25	This substitutes the exact form of $h_\theta(x^{(i)})$ used by logistic regression into the gradient descent update.
☑ $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)} \right) x^{(i)}$.	✔	0.25	This is a vectorized version of gradient descent that substitues in the exact form of $h_\theta(x^{(i)})$ used by logistic regression.
☐ $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\theta^T x - y^{(i)} \right) x^{(i)}$.	✔	0.25	This vectorized version uses the linear regression hypothesis $\theta^T x$ instead of that for logistic regression.
☐ $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\theta^T x - y^{(i)} \right) x_j^{(i)}$ (simultaneously update for all j).	✔	0.25	This uses the linear regression hypothesis $\theta^T x$ instead of that for logistic regression.
Total		1.00 / 1.00	

Question 5

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

Your Answer		Score	Explanation
☐ Since we train one classifier when there are two classes, we train two classifiers when there are three classes (and we do one-vs-all classification).	✔	0.25	We need to train three classifiers if there are three classes; each one treats one of the three classes as the $y = 1$ examples and the rest as the $y = 0$ examples.
☐ The cost function $J(\theta)$ for logistic regression trained with $m \geq 1$ examples is always greater than or equal to zero.	✖	0.00	The cost for any example $x^{(i)}$ is always ≥ 0 since it is the negative log of a quantity less than one. The cost function $J(\theta)$ is a summation over the cost for each eample, so the cost function itself must be greater than or equal to zero.
☑ The one-vs-all technique allows you to use logistic regression for problems in which each $y^{(i)}$ comes from a fixed, discrete set of values.	✔	0.25	If each $y^{(i)}$ is one of k different values, we can give a label to each $y^{(i)} \in \{1, 2, \ldots, k\}$ and use one-vs-all as described in the lecture.
☑ Linear regression always works well for classification if you classify by using a threshold on the prediction made by linear regression.	✖	0.00	As demonstrated in the lecture, linear regression often classifies poorly since its training prodcedure focuses on predicting real-valued outputs, not classification.
Total		0.50 /	

1.00