Feedback — VI. Logistic Regression

Help Center

You submitted this quiz on **Sat 14 Feb 2015 1:29 PM CET**. You got a score of **5.00** out of **5.00**.

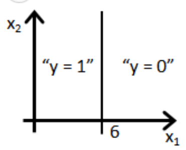
Question 1

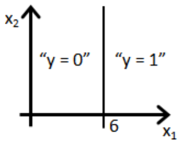
Suppose that you have trained a logistic regression classifier, and it outputs on a new example x a prediction $h_\theta(x)$ = 0.7. This means (check all that apply):

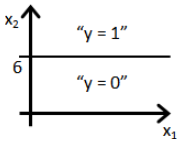
Your Answer		Score	Explanation
☐ Our estimate for $P(y = 0\|x; \theta)$ is 0.7.	✔	0.25	$h_\theta(x)$ is $P(y = 1\|x; \theta)$, not $P(y = 0\|x; \theta)$.
☑ Our estimate for $P(y = 1\|x; \theta)$ is 0.7.	✔	0.25	$h_\theta(x)$ is precisely $P(y = 1\|x; \theta)$, so each is 0.7.
☑ Our estimate for $P(y =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

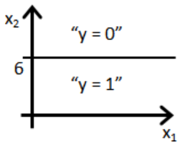
Question 2

Suppose you train a logistic classifier $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$. 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.





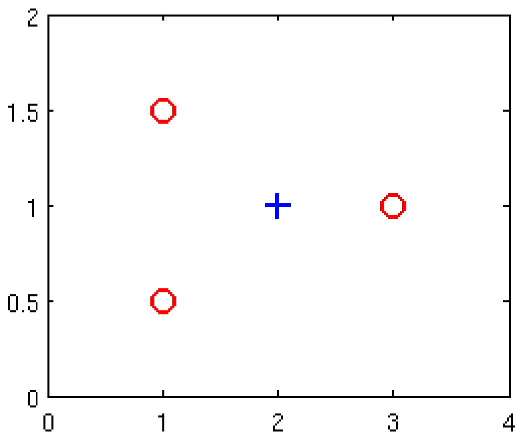


Total	1.00 / 1.00

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
$J(\theta)$ will be a convex function, so gradient descent should converge to the global minimum.	✔	0.25	The cost function $J(\theta)$ is guaranteed to be convex for logistic regression.
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.
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.
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.
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
$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\theta^T x - y^{(i)}\right) x_j^{(i)}$	✔	0.25	This uses the linear regression

(simultaneously update for all j).			hypothesis $\theta^T x$ instead of that for logistic regression.
☑ $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)} \right) x^{(i)}$.	✔	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.
☑ $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)} \right) x_j^{(i)}$ (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.
Total		1.00 / 1.00	

Question 5

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

Your Answer		Score	Explanation
☑ The cost function $J(\theta)$ for logistic regression trained with $m \geq 1$ examples is always greater than or equal to zero.	✔	0.25	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.
☐ For logistic regression, sometimes gradient descent will converge to a local minimum (and fail to find the global minimum). This is the reason we prefer more advanced optimization	✔	0.25	The cost function for logistic regression is convex, so gradient descent will always converge to the global minimum. We still might use a more advanded optimization algorithm since they can be faster and don't require you to select a learning rate.

algorithms such as fminunc (conjugate gradient/BFGS/L-BFGS/etc).			
☐ Linear regression always works well for classification if you classify by using a threshold on the prediction made by linear regression.	✔	0.25	As demonstrated in the lecture, linear regression often classifies poorly since its training prodcedure focuses on predicting real-valued outputs, not classification.
☑ The sigmoid function $g(z) = \frac{1}{1+e^{-z}}$ is never greater than one (> 1).	✔	0.25	The denomiator ranges from ∞ to 1 as z grows, so the result is always in $(0, 1)$.
Total		1.00 / 1.00	