# Feedback — VI. Logistic Regression

Help

You submitted this quiz on **Mon 31 Mar 2014 12:03 PM PDT**. You got a score of **2.00** out of **5.00**. You can attempt again in 10 minutes.

## Question 1

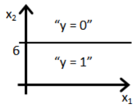
Suppose that you have trained a logistic regression classifier, and it outputs on a new example $x$ a prediction $h_\theta(x)$ = 0.4. This means (check all that apply):

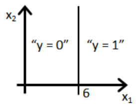
| Your Answer | | Score | Explanation |
|---|---|---|---|
| ☑ Our estimate for $P(y = 1\|x; \theta)$ is 0.4. | ✔ | 0.25 | $h_\theta(x)$ is precisely $P(y = 1\|x; \theta)$, so each is 0.4. |
| ☐ Our estimate for $P(y = 0\|x; \theta)$ is 0.4. | ✔ | 0.25 | $h_\theta(x)$ is $P(y = 1\|x; \theta)$, not $P(y = 0\|x; \theta)$ |
| ☑ Our estimate for $P(y = 0\|x; \theta)$ is 0.6. | ✔ | 0.25 | Since we must have $P(y = 0\|x; \theta) = 1 - P(y = 1\|x; \theta)$, the former is 1 - 0.4 = 0.6. |
| ☐ Our estimate for $P(y = 1\|x; \theta)$ is 0.6. | ✔ | 0.25 | $h_\theta(x)$ gives $P(y = 1\|x; \theta)$, not $1 - P(y = 1\|x; \theta)$. |
| Total | | 1.00 / 1.00 | |

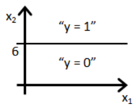
## Question 2

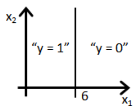
Suppose you train a logistic classifier $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$. Suppose $\theta_0 = -6, \theta_1 = 0, \theta_2 = 1$. Which of the following figures represents the decision boundary found by your classifier?

| Your Answer | Score | Explanation |
|---|---|---|




| | | | |
|---|---|---|---|
|   | ✖ | 0.00 | In this figure, we transition from negative to positive when $x_1$ goes from below 6 to above 6, but for the given values of $\theta$, the transition occurs when $x_2$ goes from below 6 to above 6 |



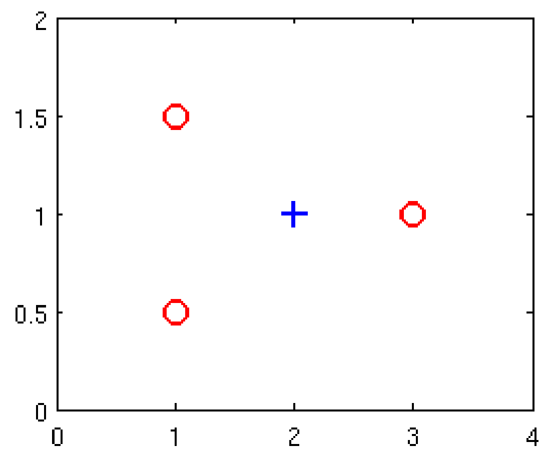




Total 0.00 / 1.00

# 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 |
|---|---|---|---|
| ☐ At the optimal value of $\theta$ (e.g., found by fminunc), we will have $J(\theta) \geq 0$ | ✖ | 0.00 | The cost function $J(\theta)$ is always non-negative for logistic regression. |

will have $J(\theta) \geq 0$.

| | | | |
|---|---|---|---|
| ☐ If we train gradient descent for enough iterations, for some examples $x^{(i)}$ in the training set it is possible to obtain $h_\theta(x^{(i)}) > 1$. | ✔ | 0.25 | The function $g(z)$ in the hypothesis $h_\theta(x)$ is the sigmoid function $\frac{1}{1+e^{-z}}$ which always lies between 0 and 1. |
| ☐ $J(\theta)$ will be a convex function, so gradient descent should converge to the global minimum. | ✖ | 0.00 | The cost function $J(\theta)$ is guaranteed to be convex for logistic regression. |
| ☑ The positive and negative examples cannot be separated using a straight line. So, gradient descent will fail to converge. | ✖ | 0.00 | 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 | | 0.25 / 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 $\alpha$? Check all that apply.

| Your Answer | | Score | Explanation |
|---|---|---|---|
| ☐ $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})x^{(i)}$. | ✖ | 0.00 | This is a vectorized version of the direct substitution of $\frac{\partial}{\partial\theta_j} J(\theta)$ into the gradient descent update. |
| ☑ $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})x^{(i)}$ (simultaneously update for all $j$). | ✖ | 0.00 | This incorrectly multiplies by the vector $x^{(i)}$ in the summation rather than just $x_j^{(i)}$. |

| | | | |
|---|---|---|---|
| $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m \left( \frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)} \right) x_j^{(i)}$ (simultaneously update for all $j$). | ✖ | 0.00 | This substitutes the exact form of $h_\theta(x^{(i)})$ used by logistic regression into the gradient descent update. |
| $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^m \left( \theta^T x - y^{(i)} \right) x^{(i)}$. | ✔ | 0.25 | This vectorized version uses the linear regression hypothesis $\theta^T x$ instead of that for logistic regression. |
| Total | | 0.25 / 1.00 | |

# Question 5

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

| Your Answer | | Score | Explanation |
|---|---|---|---|
| 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 $\infty$ to 1 as $z$ grows, so the result is always in $(0, 1)$. |
| For logistic regression, sometimes gradient descent will converge to a local minimum (and fail to find the global minimum). This is | ✖ | 0.00 | 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. |

| | | | |
|---|---|---|---|
| minimum). This is the reason we prefer more advanced optimization algorithms such as fminunc (conjugate gradient/BFGS/L-BFGS/etc). | | | |
| 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. | ✖ | 0.00 | 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. |
| Total | | 0.50 / 1.00 | |