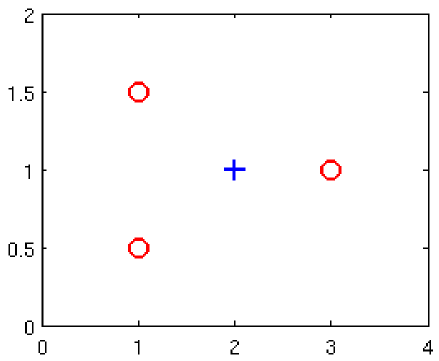
1. Suppose that you have trained a logistic regression classifier, and it outputs on a new example x a prediction $h_\theta(x)$ = 0.4. This means (check all that apply):

- [ ] Our estimate for $P(y=0|x;\theta)$ is 0.4.
- [ ] Our estimate for $P(y=1|x;\theta)$ is 0.6.
- [x] Our estimate for $P(y=1|x;\theta)$ is 0.4.
- [x] Our estimate for $P(y=0|x;\theta)$ is 0.6.

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



2 wrong

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

- [ ] $J(\theta)$ will be a convex function, so gradient descent should converge to the global minimum.
- [x] 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.
- [x] The positive and negative examples cannot be separated using a straight line. So, gradient descent will fail to converge.
- [ ] 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.

2 try below right

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

- [x] $J(\theta)$ will be a convex function, so gradient descent should converge to the global minimum.
- [x] 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.
- [ ] The positive and negative examples cannot be separated using a straight line. So, gradient descent will fail to converge.
- [ ] 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.

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

- [ ] $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\theta^T x - y^{(i)}\right) x^{(i)}$.
- [x] $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$ (simultaneously update for all j).
- [x] $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)}\right) x_j^{(i)}$ (simultaneously update for all j).
- [ ] $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}$ (simultaneously update for all j).

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

- [x] $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{1+e^{-\theta^T x^{(i)}}} - y^{(i)}\right) x^{(i)}$.
- [ ] $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}$.
- [x] $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\theta^T x - y^{(i)}\right) x_j^{(i)}$ (simultaneously update for all j).
- [ ] $\theta := \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\theta^T x - y^{(i)}\right) x^{(i)}$.

above 3 wrong

3wrong

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

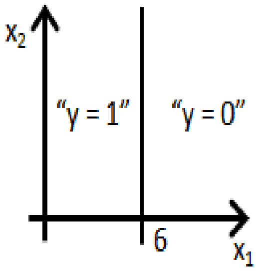
- [x] The cost function $J(\theta)$ for logistic regression trained with $m \geq 1$ examples is always greater than or equal to zero.
- [ ] Linear regression always works well for classification if you classify by using a threshold on the prediction made by linear regression.
- [ ] For logistic regression, 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 algorithms such as fminunc (conjugate gradient/BFGS/L-BFGS/etc).
- [x] The sigmoid function $g(z) = \frac{1}{1+e^{-z}}$ is never greater than one (> 1).

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

- [x] 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.
- [ ] 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 algorithms such as fminunc (conjugate gradient/BFGS/L-BFGS/etc).
- [ ] 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).
- [x] The cost function $J(\theta)$ for logistic regression trained with $m \geq 1$ examples is always greater than or equal to zero.

5. 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?

- Figure:



5. 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?

- Figure: