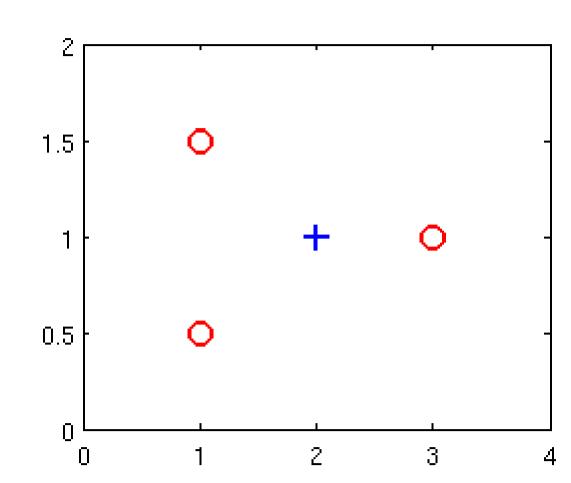
point

- 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):
  - Our estimate for  $P(y=0|x;\theta)$  is 0.3.
  - Our estimate for  $P(y=1|x;\theta)$  is 0.3.
  - Our estimate for P(y=0|x; heta) is 0.7.
  - Our estimate for  $P(y=1|x;\theta)$  is 0.7.

1 point 2. Suppose you have the following training set, and fit a logistic regression classifier  $h_{\theta}(x)=g(\theta_0+\theta_1x_1+\theta_2x_2).$ 

	$x_1$	$x_2$	у
	1	0.5	0
	1	1.5	0
	2	1	1
	3	1	0



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

- Adding polynomial features (e.g., instead using  $h_\theta(x)=g(\theta_0+\theta_1x_1+\theta_2x_2+\theta_3x_1^2+\theta_4x_1x_2+\theta_5x_2^2) \text{ ) could increase how well we can fit the training data.}$ 
  - At the optimal value of heta (e.g., found by fminunc), we will have  $J( heta) \geq 0$ .
- Adding polynomial features (e.g., instead using  $h_{ heta}(x)=g( heta_0+ heta_1x_1+ heta_2x_2+ heta_3x_1^2+ heta_4x_1x_2+ heta_5x_2^2)$  ) would increase J( heta) because we are now summing over more terms.
  - If we train gradient descent for enough iterations, for some examples  $x^{(i)}$  in the training set it is possible to obtain  $h_{ heta}(x^{(i)})>1.$

1 point

- 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  $\alpha$ ? Check all that apply.
  - $heta_j := heta_j lpha rac{1}{m} \sum_{i=1}^m \left( h_ heta(x^{(i)}) y^{(i)} 
    ight) x_j^{(i)}$  (simultaneously update for all j).
  - $hilde{ hilde hilde b} hilde hil$
  - $igcap heta := heta lpha rac{1}{m} \sum_{i=1}^m \left( heta^T x y^{(i)} 
    ight) x^{(i)}.$
  - hinspace hinspaceeee hinspace hinspace hinspace hinspace hinspace hinspace

1 point

- 4. Which of the following statements are true? Check all that apply.
  - 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.
  - The sigmoid function  $g(z)=rac{1}{1+e^{-z}}$  is never greater than one (>1).
  - 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).

point

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

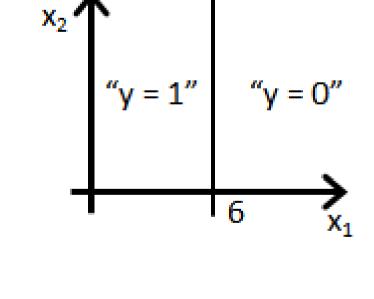


Figure:

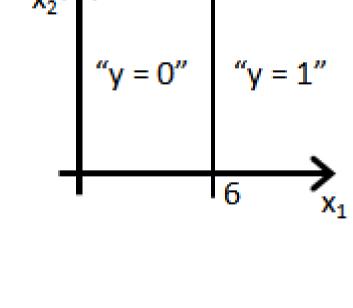


Figure:

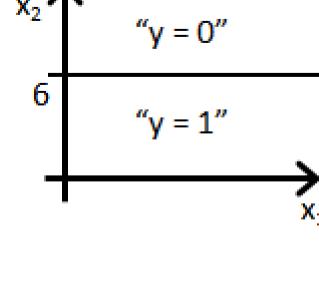
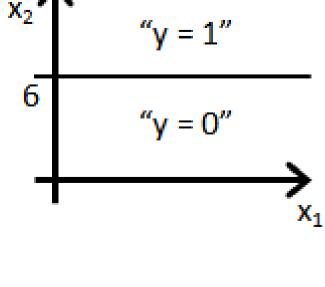


Figure:



I, Anderson Hitoshi Uyekita, understand that submitting work that isn't my own may result in permanent

failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Submit Quiz