Week 3

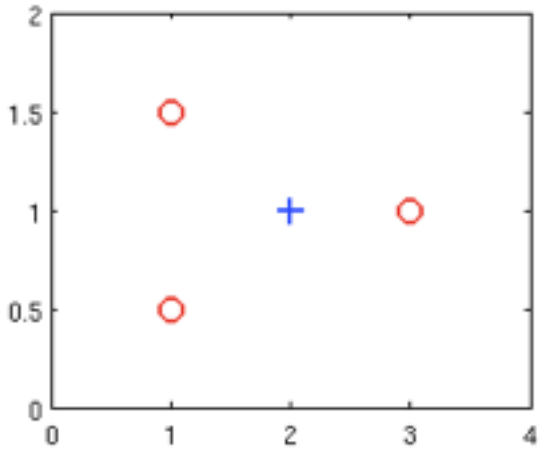
Quiz 1-Logistic Regression

Ques1. Suppose that you have trained a logistic regression classifier, and it outputs on a new example x a prediction h θ(x) = 0.7. This means (check all that apply):

Answer:
i) Our estimate for P(y=1|x;θ) is 0.7.
ii) Our estimate for P(y=0|x;θ) is 0.3.

Ques2. Suppose you have the following training set, and fit a logistic regression classifier hθ(x)=g(θ0+θ1x1+θ2x2).

x_1	x_2	y
1	0.5	0
1	1.5	0
2	1	1
3	1	0



Answer:
i) Adding polynomial features could increase how well we can fit the training data.
ii) At the optimal value of θ (e.g., found by fminunc), we will have J(θ)≥0. (As a linear decision boundary could not perfectly fit the dataset)

Ques3.

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

Ques4. Which of the following statements are true? Check all that apply.

A) The cost function J(θ) for logistic regression trained with m≥1 examples is always greater than or equal to zero.

B) 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).

C) 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).

D) 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.

Answer:
A) Correct.
B) Wrong. The reason of choosing advanced algorithms is because they don't need to choose alpha
C) Wrong. We need to train k classifier when there are k classes (k>1).
D) Correct.

Ques5. Suppose you train a logistic classifier $h\theta(x)=g(\theta o+\theta 1x1+\theta 2x2)$. Suppose $\theta o=6,\theta 1=-1,\theta 2=0$. Which of the following figures represents the decision boundary found by your classifier?

Answer:

Figure: