**1. Question 1**

Suppose that you have trained a logistic regression classifier, and it outputs on a new example xxx a prediction hθ(x)h\_\theta(x)hθ​(x) = 0.2. This means (check all that apply):



Our estimate for P(y=1∣x;θ)P(y=1|x;\theta)P(y=1∣x;θ) is 0.8.



**Our estimate for P(y=0∣x;θ)P(y=0|x;\theta)P(y=0∣x;θ) is 0.8.**



Our estimate for P(y=0∣x;θ)P(y=0|x;\theta)P(y=0∣x;θ) is 0.2.



**Our estimate for P(y=1∣x;θ)P(y=1|x;\theta)P(y=1∣x;θ) is 0.2.**

Question 2

1

Point

**2. Question 2**

Suppose you have the following training set, and fit a logistic regression classifier hθ(x)=g(θ0+θ1x1+θ2x2)h\_\theta(x) = g(\theta\_0 + \theta\_1x\_1 + \theta\_2 x\_2)hθ​(x)=g(θ0​+θ1​x1​+θ2​x2​).

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



J(θ)J(\theta)J(θ) will be a convex function, so gradient descent should converge to the global minimum.



**Adding polynomial features (e.g., instead using hθ(x)=g(θ0+θ1x1+θ2x2+θ3x12+θ4x1x2+θ5x22)h\_\theta(x) = g(\theta\_0 + \theta\_1x\_1 + \theta\_2 x\_2 + \theta\_3 x\_1^2 + \theta\_4 x\_1 x\_2 + \theta\_5 x\_2^2)hθ​(x)=g(θ0​+θ1​x1​+θ2​x2​+θ3​x12​+θ4​x1​x2​+θ5​x22​) ) 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.



At the optimal value of θ\thetaθ (e.g., found by fminunc), we will have J(θ)≥0J(\theta) \geq 0J(θ)≥0.



**Adding polynomial features (e.g., instead using hθ(x)=g(θ0+θ1x1+θ2x2+θ3x12+θ4x1x2+θ5x22)h\_\theta(x) = g(\theta\_0 + \theta\_1x\_1 + \theta\_2 x\_2 + \theta\_3 x\_1^2 + \theta\_4 x\_1 x\_2 + \theta\_5 x\_2^2)hθ​(x)=g(θ0​+θ1​x1​+θ2​x2​+θ3​x12​+θ4​x1​x2​+θ5​x22​) ) would increase J(θ)J(\theta)J(θ) because we are now summing over more terms.**



If we train gradient descent for enough iterations, for some examples x(i)x^{(i)}x(i) in the training set it is possible to obtain hθ(x(i))>1h\_\theta(x^{(i)}) > 1hθ​(x(i))>1.

Question 3

1

Point

**3. Question 3**

For logistic regression, the gradient is given by ∂∂θjJ(θ)=1m∑i=1m(hθ(x(i))−y(i))xj(i)\frac{\partial}{\partial \theta\_j} J(\theta) =\frac{1}{m}\sum\_{i=1}^m{ (h\_\theta(x^{(i)}) - y^{(i)}) x\_j^{(i)}}∂θj​∂​J(θ)=m1​∑i=1m​(hθ​(x(i))−y(i))xj(i)​. Which of these is a correct gradient descent update for logistic regression with a learning rate of α\alphaα? Check all that apply.



**θ:=θ−α1m∑i=1m(θTx−y(i))x(i) \theta := \theta - \alpha \frac{1}{m} \sum\_{i=1}^m{ \left(\theta^T x - y^{(i)}\right) x^{(i)}} θ:=θ−αm1​∑i=1m​(θTx−y(i))x(i).**



θj:=θj−α1m∑i=1m(11+e−θTx(i)−y(i))xj(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)}} θj​:=θj​−αm1​∑i=1m​(1+e−θTx(i)1​−y(i))xj(i)​ (simultaneously update for all jjj).



θj:=θj−α1m∑i=1m(hθ(x(i))−y(i))x(i)\theta\_j := \theta\_j - \alpha \frac{1}{m} \sum\_{i=1}^m{ (h\_\theta(x^{(i)}) - y^{(i)}) x^{(i)}} θj​:=θj​−αm1​∑i=1m​(hθ​(x(i))−y(i))x(i) (simultaneously update for all jjj).



**θj:=θj−α1m∑i=1m(hθ(x(i))−y(i))xj(i)\theta\_j := \theta\_j - \alpha \frac{1}{m} \sum\_{i=1}^m{ (h\_\theta(x^{(i)}) - y^{(i)}) x\_j^{(i)}} θj​:=θj​−αm1​∑i=1m​(hθ​(x(i))−y(i))xj(i)​ (simultaneously update for all jjj).**

Question 4

1

Point

**4. Question 4**

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



**The sigmoid function g(z)=11+e−zg(z) = \frac{1}{1 + e^{-z}}g(z)=1+e−z1​ is never greater than one (>1 >1 >1).**



**The cost function J(θ)J(\theta)J(θ) for logistic regression trained with m≥1m \geq 1m≥1 examples is always 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 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.

Question 5

1

Point

**5. Question 5**

Suppose you train a logistic classifier hθ(x)=g(θ0+θ1x1+θ2x2)h\_\theta(x) = g(\theta\_0 + \theta\_1x\_1 + \theta\_2 x\_2)hθ​(x)=g(θ0​+θ1​x1​+θ2​x2​). Suppose θ0=6,θ1=−1,θ2=0\theta\_0 = 6, \theta\_1 = -1, \theta\_2 = 0θ0​=6,θ1​=−1,θ2​=0. Which of the following figures represents the decision boundary found by your classifier?



Figure:



Figure:



Figure:



Figure:



Je soussigné(e), **Nicolas Xuyen Lagaillardie**, accepte que ma soumission de travaux ne résultant pas de mes propres efforts entraîne un échec permanent de ce cours et la désactivation de mon compte Coursera.