Logistic Regression

5 questions

1  
point

1.

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



Our estimate for *P*(*y*=0|*x*;*θ*) is 0.6.



Our estimate for *P*(*y*=1|*x*;*θ*) is 0.4.



Our estimate for *P*(*y*=1|*x*;*θ*) is 0.6.



Our estimate for *P*(*y*=0|*x*;*θ*) is 0.4.

1  
point

2.

Suppose you have the following training set, and fit a logistic regression classifier *hθ*(*x*)=*g*(*θ*0+*θ*1*x*1+*θ*2*x*2).

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



Adding polynomial features (e.g., instead using *hθ*(*x*)=*g*(*θ*0+*θ*1*x*1+*θ*2*x*2+*θ*3*x*21+*θ*4*x*1*x*2+*θ*5*x*22) ) could increase how well we can fit the training data.



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



Adding polynomial features (e.g., instead using *hθ*(*x*)=*g*(*θ*0+*θ*1*x*1+*θ*2*x*2+*θ*3*x*21+*θ*4*x*1*x*2+*θ*5*x*22) ) would increase *J*(*θ*)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 *hθ*(*x*(*i*))>1.

1  
point

3.

For logistic regression, the gradient is given by ∂∂*θjJ*(*θ*)=1*m*∑*mi*=1(*hθ*(*x*(*i*))−*y*(*i*))*x*(*i*)*j*. Which of these is a correct gradient descent update for logistic regression with a learning rate of *α*? Check all that apply.



*θj*:=*θj*−*α*1*m*∑*mi*=1(11+*e*−*θTx*(*i*)−*y*(*i*))*x*(*i*)*j* (simultaneously update for all *j*).



*θj*:=*θj*−*α*1*m*∑*mi*=1(*hθ*(*x*(*i*))−*y*(*i*))*x*(*i*)*j* (simultaneously update for all *j*).



*θj*:=*θj*−*α*1*m*∑*mi*=1(*hθ*(*x*(*i*))−*y*(*i*))*x*(*i*) (simultaneously update for all *j*).



*θ*:=*θ*−*α*1*m*∑*mi*=1(*θTx*−*y*(*i*))*x*(*i*).

1  
point

4.

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



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).



The cost function *J*(*θ*) for logistic regression trained with *m*≥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*)=11+*e*−*z* is never greater than one (>1).

1  
point

5.

Suppose you train a logistic classifier *hθ*(*x*)=*g*(*θ*0+*θ*1*x*1+*θ*2*x*2). Suppose *θ*0=6,*θ*1=−1,*θ*2=0. Which of the following figures represents the decision boundary found by your classifier?



Figure:



Figure:

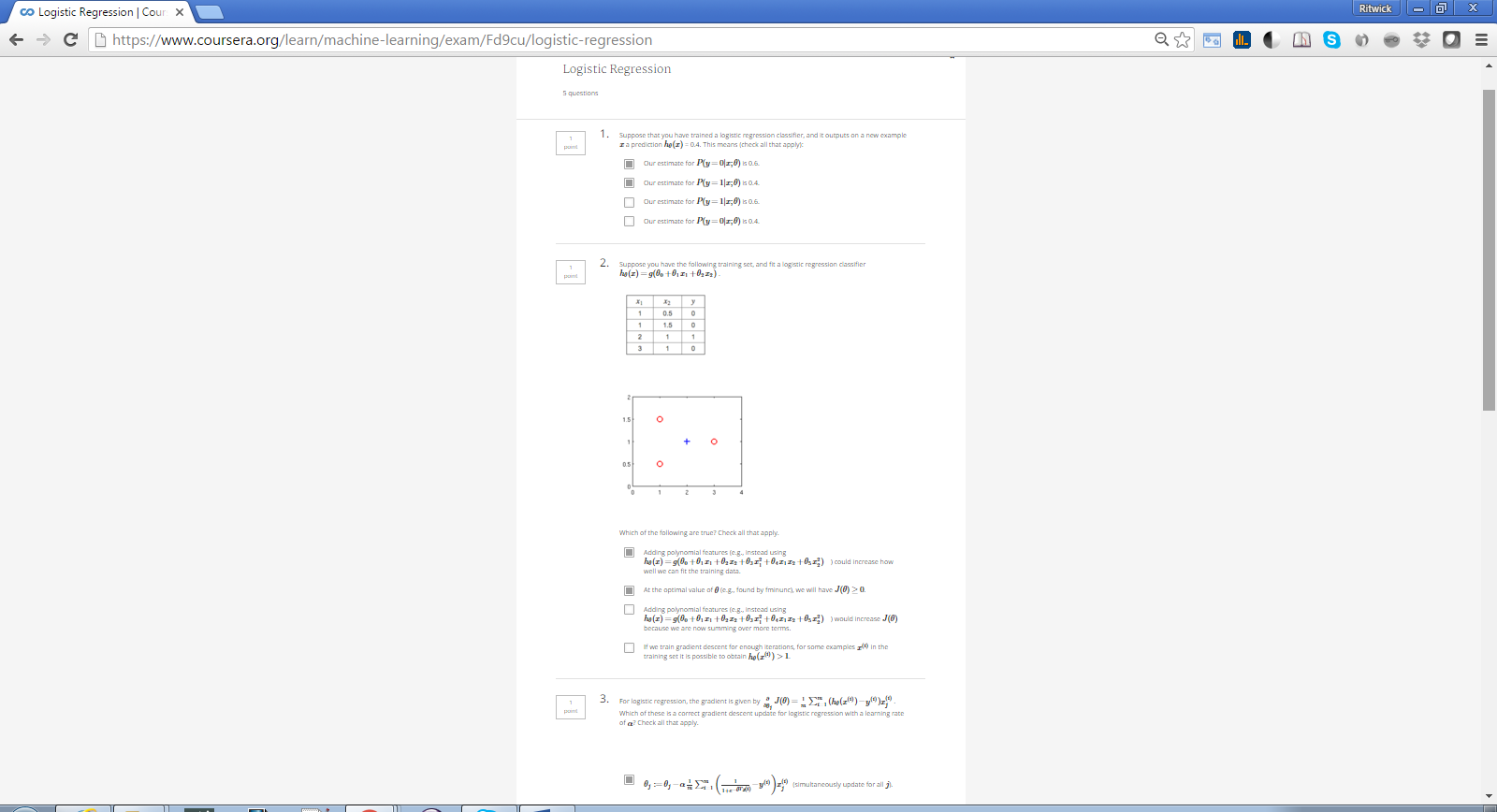


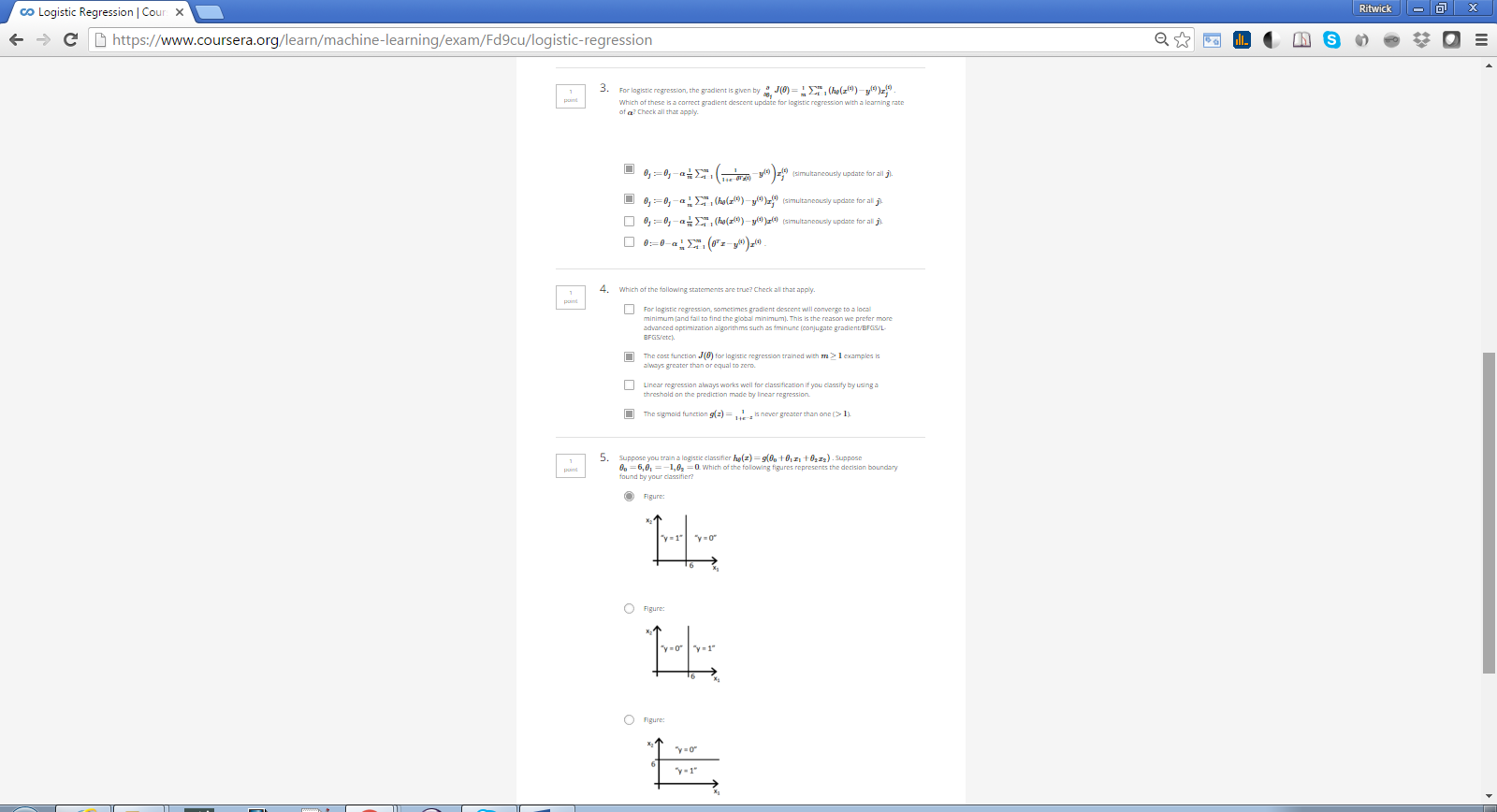
Figure:

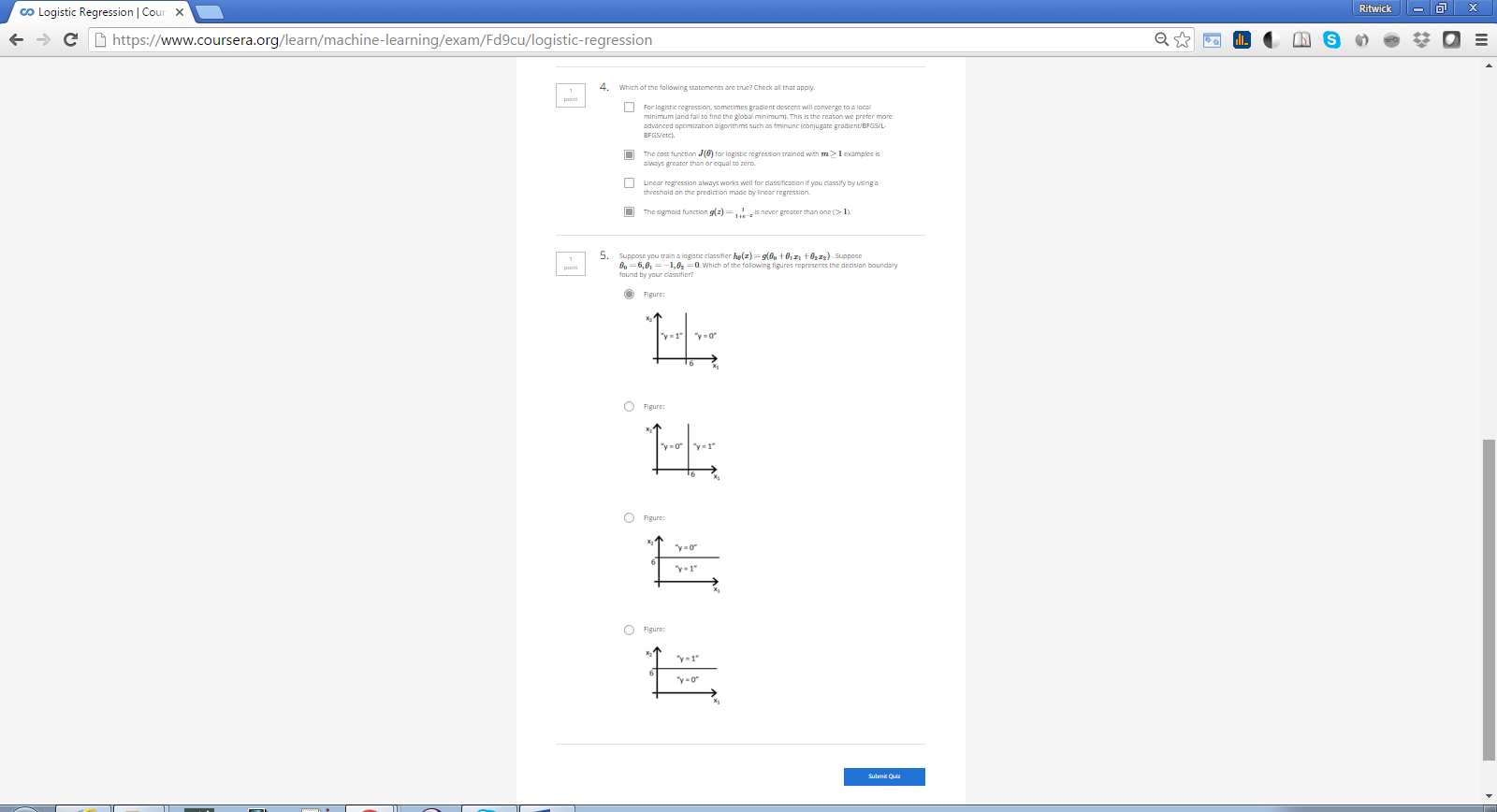


Figure:

Attempt 1







Result #1:

