CS691 Quiz 6 Name ___________________________

1. We want to find the (x_1, x_2) that minimizes the following objective function $f(x_1, x_2)$. Assuming gradient descent works, what values of x_1 and x_2 would it return? How do you know? (2 points)

$$f(x_1, x_2) = x_1^2 + x_2^2$$

df/dx = 2x1 + 2x2 = 0 x1=0 x2=0

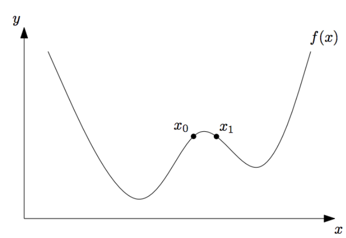
When df/dx=0 is when gradient descent would stop

2. Recall our regularized loss function. In this particular loss function, we have used a square loss $(y - \hat{y})^2$ with a $\lambda ||w||^2$ regularizer. The $\frac{1}{2}$ just helps when doing derivatives. Find ∇L_w and $\frac{\delta L}{\delta b}$. (5 points)

$$L(w, b) = \sum_n (y - (w \bullet x_n + b))^2 + \frac{\lambda}{2} ||w||^2$$

dL/dw = -2xn(y-(wxn+b)) + lambda*w
dL/db = -2(y-(wxn+b))

3. What are the possible outcomes of gradient descent on the following function? *Explain and indicate on the graph what might happen.* IGNORE THE x_0 AND x_1. THEY ARE JUST PART OF A GRAPH I STOLE FROM SOMEWHERE ELSE. (3 points)



CS491 Quiz 6 Name ___________________________

1. We want to find the (x_1, x_2) that minimizes the following objective function $f(x_1, x_2)$. Assuming gradient descent works, what values of x_1 and x_2 would it return? How do you know? (2 points)

$$f(x_1, x_2) = x_1^2 + x_2^3$$

2. Use gradient descent to minimize the following objective function. **You only need to run 3 iterations.** Show your work. Start at $x_{init} = (0, 0, 0)$, and use learning rate $\eta = 0.1$. (5 points)

$$f(x_1, x_2, x_3) = (x_1 + 1)^2 + 2x_2^3 + 10x_3$$

df/dx = (2x1+20, 6x2,10)
iter1: df(0,0,0)= (20,0,10)
x = (0,0,0) + 0.1(20,0,10) = (2,0,1)
iter2: df(2,0,1) = (24,0,10)
x = (2,0,1) + 0.1(24,0,10) = (4.4,0,2)
iter3: df(4.4,0,2) = (28.8,0,10)
x = (4.4,0,2) + 0.1(28.8,0,10) = (7.28,0,3)

3. What are the possible outcomes of gradient descent on the following function? *Explain and indicate on the graph what might happen.* IGNORE THE x_0 AND x_1. THEY ARE JUST PART OF A GRAPH I STOLE FROM SOMEWHERE ELSE. (3 points)