CS491 Quiz 6 Name Solution.

1. Run gradient descent on the function $f(x_1, x_2) = \frac{1}{2}x_1^2 + x_2^2$ starting at the point $s_0 = (2, 2)$ and using a step size of $\eta = \frac{1}{2}$. Run three steps of the algorithm. What is the output? (5 points)

$$\nabla f = \begin{bmatrix} x_1 \\ 2x_2 \end{bmatrix}$$

(1) $$\begin{matrix} x_1 \\ x_2 \end{matrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

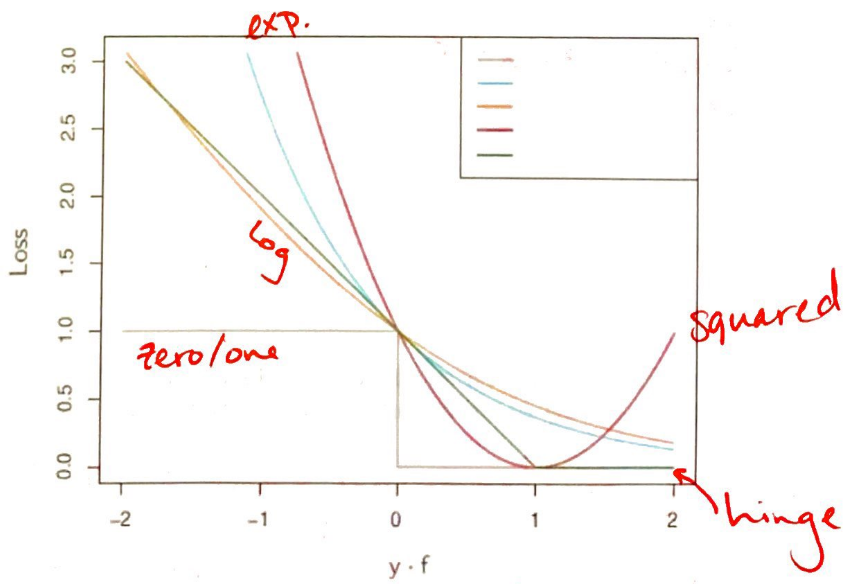
(2) $$\begin{matrix} x_1 \\ x_2 \end{matrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}$$

(3) $$\begin{matrix} x_1 \\ x_2 \end{matrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ 0 \end{bmatrix}$$

Output

$$\begin{bmatrix} \frac{1}{4} \\ 0 \end{bmatrix}$$

2. Label each of the following loss functions using these options: 0/1, hinge, sigmoid, squared, absolute, exponential, log. (5 points)



CS691 Quiz 6 Name Solution.

1. We want to find the w that minimizes the following objective function $L(w)$. What values of w might gradient descent return? Explain. (3 points)

$$L(x) = x^3$$

$w = 0$

$w = -\infty$

$\frac{dL}{dw} = 3w^2 = 0$ @ $w = 0$

But if we pass $w = 0$ gradient descent could take us down the curve forever.

2. Recall our regularized loss function. In this particular loss function, we have used an exponential loss $e^{y\hat{y}}$ with a $||w||^2$ regularizer. Find ∇L_w and $\frac{\delta L}{\delta b}$. (5 points)

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

$\frac{dL}{db} = e^{y_n \hat{y}_n} \cdot y_n$

$\nabla L_w = e^{y_n \hat{y}_n} \cdot y_n x_n + \lambda \vec{w}$

$\hat{y}_n = w \bullet x_n + b$

3. If I am performing gradient descent on the following function: $L(w) = w^2$ starting with w=3. What would cause gradient descent to not reach the minimum and return $w = 0$? (2 points)

If my η (learning rate) is too high, I will never reach the minimum