CSC311 Non-programming

David Pham: 1002597825

Oct 15 2020

Question 4 (e).
Given that $x_j^{(i)} = X_{ij}, y^{(i)} = [y]_i, t^{(i)} = [t]_i, [X^T]_{ji} = X_{ij}$ and $[\frac{\partial \mathcal{J}}{\partial w}]_j = \frac{\partial \mathcal{J}}{\partial w_j}$

$$\frac{\partial \mathcal{J}}{\partial w_j} = \sum_{i=1}^{N}(y^{(i)} - t^{(i)})x_j^{(i)}/N \iff \sum_{i=1}^{N}([y]_i - [t]_i)X_{ij}/N \iff \sum_{i=1}^{N}[y-t]_i[X^T]_{ji}/N$$

I AM LOST FROM HERE

Question 4 (f).
We are trying to prove

$$\mathcal{L}_{LCE}(z,t) = \mathcal{L}_{CE}(\sigma(z),t) = tlog(1+e^{-z}) + (1-t)log(1+e^{-z})$$

We know from the previous slide, we know that

$$\mathcal{L}_{CE}(y,t) = -tlogy - (1-t)log(1-y)$$

We know that $y = \sigma(z) = \frac{1}{1+e^{-z}}$. Therefore we get

$$\mathcal{L}_{CE}(\sigma(z),t) = -tlog(\sigma(z)) - (1-t)log(1-\sigma(z))$$

$$= -t(\frac{1}{1+e^{-z}}) - (1-t)log(1-(\frac{1}{1+e^{-z}}))$$

$$= -t(log1 - log(1+e^{-z})) - (1-t)log(\frac{1}{1+e^{-z}})$$

$$= -t(log1 - log(1+e^{-z})) - (1-t)(log1 - log(1+e^{-z}))$$

$$= -t(-log(1+e^{-z})) - (1-t)(-log(1+e^{-z}))$$

$$= tlog(1+e^{-z}) + (1-t)log(1+e^{-z})$$

as needed

Question 6(e).
The reason that the validation accuracy is higher for the digits 4 and 7 is that the clusters of the data is more split and definitive therefore when going into a higher k, it can refine the correctness. The problem with 5 and 6 is that they have tighter clusters so it has a more likely chance to pick

the wrong image as k increases. When analyzing more ambiguous data, 4 and 7 would get marked correctly while 5 and 6 would be less accurate in its result.

Question 6(f).
The reason why we choose odd values of K is that if we were to choose even values of K, you will run the risk that there exists a tie in the decision which can cause problems. Therefore in order to guarantee no ties, we choose odd values of K

Question 6(g).
With MINST data, there is not an extensive analysis on the data set. We either have the number we desire or we do not. Let us take 0 for an example, if we are trying to determine whether a number is 0, we only have to check if there exists values that in a ring. Therefore it produces high accuracies.