CS491 Quiz 8 Name Solution

1. As the amount of training data increases, how do traditional machine learning methods compare to deep learning? Explain. (2 points)

Deep learning allows for learning of the "best" features given the data & the task (label).
More data = better features & model.
Traditional ML methods don't improve features.

2. What is the result of 2x2 Max Pooling on the following input? (3 points)

5	1	4	1	9
5	1	4	1	9
5	1	4	1	9
5	1	4	1	9
5	1	1	1	9

5 4 4 9
5 4 4 9
5 4 4 9
5 4 4 9

3. I am trying to create a neural network to predict whether a word is a noun or a verb. Should I use a softmax, sigmoid cross-entropy, or euclidean loss? Explain. (2 points)

Any answer w/ reasonable justification was accepted.

4. Consider a CNN with an input image of size 10x10. The output of the network is 8x8. The network has only one layer. Assume no zero padding, and a stride of 1. How many parameters are there in a CNN with this input and output? Explain. (3 points)

10x10 → 8x8 means 3x3 filter.

3x3 = 9 weights

CS691 Quiz 8 Name Solution

1. Given an input image of size NxN, filter size of fxf, stride of s, and zero padding of z pixels in each direction, give an expression in terms of these variables for the size of the first hidden layer. (3 points)

$$\frac{N+2z-f}{s}+1$$

2. Given an input image of 25x25x3, that is, a color image of size 25x25, what size filter is needed to get a first hidden layer of size 21x21? Assume no zero padding, and stride=1. (2 points)

5x5x3

3. What is weight decay? How does it relate to our previous optimization problem of minimizing L with respect to w? Remember one example of L was $L(y,\hat{y}) = \frac{1}{2}(y-\hat{y})^2 + \frac{\lambda}{2}||w||^2$. (2 points)

reducing the value of the weights by some small factor.
Similar to the regularization $\frac{\lambda}{2}||w||^2$.

4. What is the result of 2x2 Max Pooling on the following input? (3 points)

5	1	4	1	9
5	1	4	1	9
5	1	4	1	9
5	1	4	1	9
5	1	1	1	9

5 4 4 9
5 4 4 9
5 4 4 9
5 4 4 9