MCL 606

Deep Learning

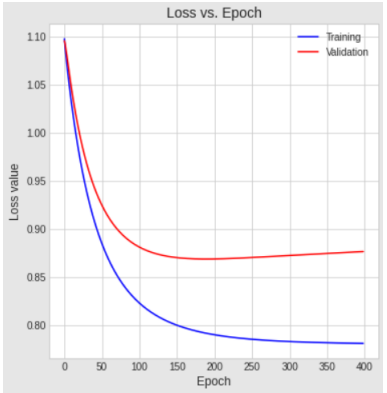
Even Sem. 2021

Sessional-1 — April 23, 2021— 20 Points

True/false questions need no explanation but carry negative points for incorrect answers.

1. True/false: Softmax and SVM algorithms based on a loss-function formulation we saw in class are *linear classifiers*.
2. Suppose we have 10^4 samples corresponding to 10 output labels and that each sample is a 32×32 grayscale image. If we apply the softmax algorithm to this data, what are the dimensions of the following quantities assuming that the bias-trick pre-processing has been performed?
   - Data matrix.
   - Weight matrix.
   - Probability matrix.
   - Adjusted probability matrix.
   - Total average data loss.
   - Regularization loss.
   - Gradient of total loss w.r.t. the weight matrix.
3. In the setup of the previous question, suppose we
   - randomly initialize the "weights part" of the weight matrix,
   - initialize the "bias part" of the weight matrix to zeros.

   The total average data loss calculated using the resulting weight matrix is approximately equal to _____.
4. True/false: Suppose we initialized the entire weight matrix to zeros and performed several iterations of gradient descent. The weight parameters in the "weight part" of the weight matrix will remain as zeros after update at every step.
5. Using the same dimensions given in Question-2, suppose that for training you apply gradient descent using batch processing with batch size equal to 20. How many times will the weight matrix be updated every epoch?
6. Calculate the first gradient below. For the remaining, just state the dimensions of the gradient object:
   - $\nabla_w \left(\sqrt{\|w\|^2 + 1} \right)$, where w is an n-vector.
   - $\nabla_w \left(\frac{w}{\sqrt{\|w\|^2+1}} \right)$, where w is an n-vector.
   - $\nabla_W (Wa)$, where W is a 3×4-matrix and a is a 4-vector.
7. True/false: The derivative of the loss with respect to some weight parameter evaluated at its current value is -10. That means that decreasing this weight (by a tiny amount) would decrease the loss.
8. Suppose you train the SVM algorithm on a data set resulting in the following plot:



   In just one line, explain at what epoch you would stop the training and why.