

HW4 Written Part

Due 5pm Friday Feb 28, 2020

What is your name?

Problem 1

Suppose that you have a classification model that provides about 80% classification accuracy on the training set and on out-of-sample test data. If you ask a human to do this task, they will generally be able to achieve classification accuracy of close to 100%.

(a) For each of the following possible next steps in your analysis, choose one option that you might consider next, and write 1 sentence explaining your reasoning.

- Change L_2 penalty by:

Increasing λ

Decreasing λ

Why?

- Change the dropout rate by:

Increasing dropout rate

Decreasing dropout rate

Why?

- Change training time by:

Increasing the number of epochs

Decreasing the number of epochs

Why?

- Change model structure by:

Adding more hidden layers

Reducing the number of hidden layers

Why?

- Change model structure by:

Adding more units to hidden layers

Reducing the number of units in hidden layers

Why?

(b) Collecting more data is always good. In this case, would you say that collecting more data is a higher priority than trying some of the ideas in part (a), or a lower priority? Why?

Problem 2

(a) I divided my labeled classification data into a training set used for model construction and another portion for validation. I then evaluated 1000 neural architectures by estimating parameters with gradient descent on the training set and computing classification accuracy on the validation set. Discuss why the resulting model is likely to yield poorer accuracy on out-of-sample test data as compared to the validation data, even though the validation data was not directly used for estimating the bias and weight parameters. This could be answered in 1-3 sentences.

(b) Suppose I now look at performance on my test set and find that indeed, classification accuracy is lower on the test set than it was on the training and validation sets. I now continue to refine my model based on combined validation and test set performance, fitting another 200 variations on neural network models. Is my test set performance a reliable measure of the quality of my final model? This could be answered in 1 sentence.

Problem 3

(a) What problem(s) result from having a learning rate that is too high? How would you detect such problems in a plot of training set loss as a function of training epoch? You can answer in 1-2 sentences.

(a) What problem(s) result from having a learning rate that is too small? How would you detect such problems in a plot of training set loss as a function of training epoch? You can answer in 1-2 sentences.

Problem 4

I'm fitting a neural network using L_1 regularization. Suppose that in the current iteration of gradient descent, my weight parameter $w_{11}^{[1]}$ is positive. (The point of this is that although the derivative of the absolute value is not defined everywhere, it is defined if the argument to the absolute value is not 0.)

(a) Find an expression for $\frac{\partial J(b,w)}{\partial w_{11}^{[1]}}$. Your answer can involve a term like $-\frac{1}{m} \frac{\partial \ell(b,w)}{\partial w_{11}^{[1]}}$, where $\ell(b,w)$ is the log-likelihood for your model. (You should not be doing any involved chain rule calculations.)

(b) Suppose the current value of $w_{11}^{[1]}$ is small (positive, but close to 0). Is the effect of the penalty on a gradient descent update step for $w_{11}^{[1]}$ larger for L_1 regularization or L_2 regularization?

(c) In general, would you expect the final estimate of $w_{11}^{[1]}$ to be closer to 0 if you use L_1 regularization or L_2 regularization? You can answer in 1 sentence, but you should justify your answer based on your result from part (b).