

STAT 542: Homework 4

Due: Mar. 22 midnight on Canvas

Please make sure that your solutions are readable and the file size is reasonable. Typing the answers is highly encouraged.

Problem 1.

[2pts] Derive the forward and backward propagation equations for the cross-entropy loss function.

Hint: The gradient update equations for the sum of squares loss can be found in p396 of Elements of Statistical Learning, which is used for regression tasks. Here we consider the classification task and replace the sum of squares loss by the cross-entropy loss. You should highlight in your solution what parts have changed.

Problem 2.

[2pts] In a K -class classification problem, suppose that for given neural network parameter (weights) θ and input x , the output of the network are functions $f_k(x, \theta)$, $k = 1, \dots$, which are nonnegative numbers that sum to 1. Show that minimizing the cross-entropy loss is equivalent to maximum likelihood estimation, if $f_k(x, \theta)$ is interpreted as the likelihood $P(Y = k|x, \theta)$.

Problem 3.

[3pts] In the “comparison of test errors” slide we saw that the test error does not decrease further as the number of bagged trees exceeds a certain point. Give an explicit calculation to verify this in the example of linear statistics.

Hint: see exercise 15.4 in Elements of Statistical Learning.