

Machine Learning - Lecture 14

Knowledge test

Chris Thornton

November 17, 2011

Question

In the normal arrangement for a multi-layer perceptron (MLP), each hidden unit takes input from every input unit, and each output unit takes input from every hidden unit. Let's say an MLP with this architecture has three input units, three hidden units and two output units. Assuming unit bias is implemented as a weight, how many numbers do we need to specify the weights of the network? Is it (a) 5, (b) 8, (c) 9, (d) 15, (e) 18, (f) 20, (g) 24, (h) 28) or (i) 32?

Question

While training an MLP, it is noticed that training error is swinging wildly between low and high values. Which of the following would be plausible remedies? (a) decrease the learning rate, (b) increase the learning rate, (c) decrease momentum, (d) increase momentum, (e) decrease learning rate but increase momentum, (f) increase learning but decrease momentum.

Question

When using a reference point to define a linear boundary, we normally have to set a threshold value. The boundary then separates off all datapoints whose inner product with the reference point is above the threshold. In the MLP method, multiple reference points are combined to produce a curving boundary. But no thresholds are involved. What then defines the difference between one side of the boundary and the other?

Question

Cross-validation applied to a fully trained-up MLP reveals generalization error to be considerably worse than training error. It seems over-fitting has occurred. The decision is made to re-run the learning using a stronger bias. Which of the following changes might produce the desired effect? (a) increase the learning rate, (b) increase the number of hidden units, (c) decrease the number of hidden units, (e) increase the activation of the bias unit, (f) decrease the activation of the bias unit, (g) increase the number of training examples?

Question

Let's say an MLP is restricted so that unit activation is always thresholded at the 0.5 level. This means unit activation is always either precisely 0 or precisely 1. Does this have the effect of (a) strengthening the bias of the learning method, (b) weakening the bias of the method, (c) leaving the bias unchanged?

Question

A certain MLP is made up of a number of input, hidden and output units. We would like to implement unit bias. Which of the units in the network can benefit from having a bias value? How is this bias value more easily implemented?

How can the problem of over-fitting arise in the case of k-means clustering?

Question

How does the problem of over-fitting relate to the problem of lookup tables?

Question

What is left open in the the stopping condition for delta-rule error-correction? How could we formulate a specific condition for a particular domain?

Question

Is it possible to achieve delta-rule error-correction through subtraction rather than addition of error values? How would this be done?