Stats 315B: Homework 3

Joe Higgins, Austin Wang, Jessica Wetstone

Due 5/20/2018

Question 1

Consider a multi-hidden layer neural network trained by sequential steepest-descent using the weight updating formula $w_t = w_{t-1} - \eta G(w_{t-1})$ Here t labels the observations presented in sequence (time) and $G(w)$ is the gradient of the squared-error criterion evaluated at w. Derive a recursive "back-propagation" algorithm for updating all of the network weights at each step. With this algorithm the update for an input weight to a particular hidden node is computed using only the value of its corresponding input (that it weights), the value of the output of the hidden node to which it is input, and an "error signal" from each of the nodes in the next higher layer to which this node is connected. Thus, each node in the network can update its input weights using information provided only by the nodes to which it is connected.

Question 2

Consider a radial basis function network with spherical Gaussian basis of the form $B(x|\mu_m, \sigma_m) = \left(-\frac{1}{2\sigma_m^2}\sum_{j=1}^{n}(x_j - \mu_{jm})^2\right)$, with the function approximation given by $\hat{F}(x) = \sum_{m=1}^{M} a_m B(x|\mu_m, \sigma_m)$ and sum-of-squares error criterion. Derive expressions for the gradient $G(x)$ with respect to all (types of) parameters in the network.

Question 3

Question 4

Question 5

Describe K—fold cross-validation. What is it used for. What are the advan- tages/disadvantages of using more folds (increasing K). When does cross—validation estimate the performance of the actual predicting function being used.

Question 6

Suppose there are several outcome variables $\{y_1, y_2, ..., y_M\}$ associated with a common set of predictor variables $x = \{x_1, x_2, ..., x_n\}$. One could train separate single output neural networks for each outcome y_m or train a single network with multiple outputs, one for each y_m. What are the relative advantages/disadvantages of these two respective approaches. In what situations would one expect each to be better than the other.

Question 7

Spam Email. The data sets spam_stats315B_train.csv, spam_stats315B_test.csv and documentation for this problem are the same as in Homework 2 and can be found in the class web page. You need first to standardize predictors and choose all the weights starting values at random in the interval [-0.5, 0.5].