Foundations of machine learning: Assignment 1

The task set was to develop a model that takes a data set of a 5-dimensional input space and outputs a 1-dimensional continuous value. For training, a dataset of 1000 examples was given. To build the model for this regression problem a Radial Basis Function Network (RBF) was chosen over an Multi-Layer Perceptron (MLP). This was because they are quite robust to noise and they are fast to train due to how few parameters that need to be optimized in contrast to an MLP.

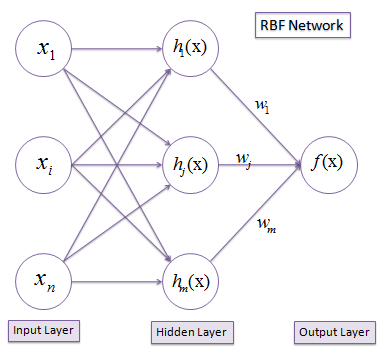
**The idea behind RBFs is being able classify points by characterising data points by the distances from basis or centre vectors chosen when using training data. RBF’s have a fixed three-layer architecture that works on Covers theorem on the separability of patterns (Cover, 1965).

Figure 1 (Saedsayad.com, 2017)

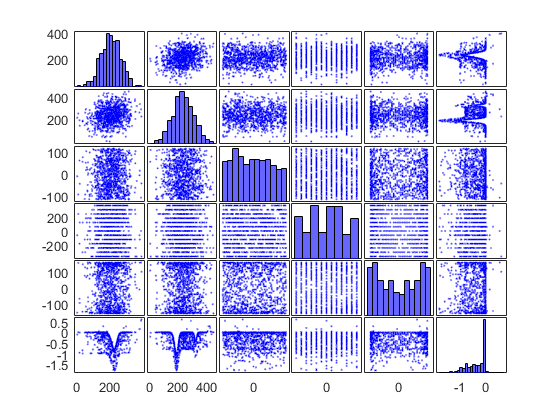
*When casting a complex pattern-classification problem, into a high dimensional space nonlinearly, it is more likely to be linearly separable than in a low-dimensional space.*

Initially the data is passed through a function, that maps the data into a layer of hidden nodes, consisting of more dimensions. Then similar in the way an MLP works the output of each node is multiplied by the connected weight as it is passed through a final function mapping to a single node

Pre-processing Techniques Used

Before designing the network to approximate the function for the given data, a few pre-processing techniques were first applied. Normalization was used first, to prevent the network from being ill-conditioned. Normalizing the inputs sets the range of each dimensional input to have roughly the same range. Ultimately allowing for a more stable convergence of weights. The equation used below:

Second was Principle Component Analysis (PCA). PCA is a dimensionality reduction technique that converts a set of ‘possibly’ correlated data to a lower dimensions. To apply PCA it was vital that the number of dimensions to reduce to was such that upwards of 95% of the variance could be retained (Ng, 2017). This can be calculated after finding the eigenvectors and eigenvalues of each dimension, one can do a summation of the retained eigenvalues over all the eigenvalues to establish this variance retention.

Furthermore, in applying PCA the covariance matrix needed to be calculated. Through this by simple observation it was clear that the two most correlated inputs were dimensions 1 and 2, which could also be seen when plotting all dimensions against each other

Ultimately, the decision was made to use normalization, feature selection and the use regularization function on a Linear RBF.

This meant that the only free parameter that needed to be used was λ the regularization parameter. Otherwise the centers were determined by the exact interperpolation matrix generated by the testing set.

The optimization of the problem then came down to a brute force approach, that lowered λ by a factor of three with every iteration over 30 iteration, testing the model using a 5-fold cross validation. As the starting from a lamb

Bibliography

Cover, T. (1965). *Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition*. 1st ed. [S.l.]: [s.n.], pp.326-334.

Ng, A. (2017). *PCA - Ufldl*. [online] Deeplearning.stanford.edu. Available at: http://deeplearning.stanford.edu/wiki/index.php/PCA [Accessed 23 Mar. 2017].

Saedsayad.com. (2017). *Radial Basis Network*. [online] Available at: http://www.saedsayad.com/artificial\_neural\_network\_rbf.htm [Accessed 23 Mar. 2017].