Assignment 4 ML

April 18, 2022

J L Gouws

19G4436

1. The weights from the first layer to the second layer(first hidden layer) are supposed to represent the contribution of the pixels to little line segments. The sencond layer is supposed to pick up on small “edges" or line segments.

2. The weights from the second layer to the third layer represent collection of the “edges" making up larger structures. These weights are supposed to represent the contribution of the small line segments to lines and circles. The weights for input into the final layer are supposed to represent how the lines and circles get added together to form the numbers.

3. The weights are not coded or worked out by hand to represent this. A computer works out the weights. The network needs some starting values for the weights, these are set randomly for ease of use. Since the network starts randomly there is no reason for the training network to follow a human’s semantic structure. The learning tries to find patterns that are natural for the network to recognize, which might not be natural/sensical to humans.

1. Machine learning is the study of finding or making a function f that takes input x(feature vector) from some set D and gives output y, where nothing is known about the relationship between x and y a priori. Supervised learning is the case where y is known for some subset of D. Unsupervised learning is the case where nothing is known about the ouput of the function. An example would be taking a picture, x, of a dog or cat and finding a function, f, to classify the picture as a picture of a dog or cat, y.

2. (a) Feature scaling is changing the values of (numeric)features so that they conform to some restriction, without significantly changing the information encoded within the features.

   (b) Distance based learning algorithms.

   - K Nearest Neighbors
   - Neural networks

   (c) Tree based algorithms

   - Random forest.
   - Random forest.

3. (a)

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5

   (b) **Advantages** It gives a better estimation of the average generalization of the model.

   Disadvantages It takes a long time to run. This is a less significant problem, but more data is generally needed.

4. (a) The gradient descent method might only find a local minimum and not the global minimum.

   (b) By looking at the error data, overfitting of the model can be seen. If the model is overfitting then the model is too complex, a lower degree polynomial is required. If the model is underfitting the model is too simple and a higher order polynomial is required.

   (c) By looking at the error data, overfitting of the model can be seen. If the model is overfitting then more regularization is needed. If the model is starting to underfit then less regularization is needed.