Machine Learning

Homework 9 solution - Shyam Arumugaswamy

1.

In neural networks, we use basis functions to map the sample into a space where it is linearly separable. For linear separability, each dataset require different transformation which is more difficult for high dimensional data. Basis functions are used as activation functions, function approximations, classification.

2.

Lets assume a two-layer neural network with only one linear output unit and sigmoid hidden units. If we change the hidden units to tanh, we must adopt the weights in the second layer. The weights in the first layer simply must be halved to construct a network equivalent to the original one. This can be extended straight forward to a network with multiple outputs and to a network with sigmoid output(s) (which also could be substituted by tanh units).

3.

$\sigma(x) = \frac{1}{1+ e^{-x}}$, $\tanh(x) = \frac{e^{x}- e^{-x}}{e^{x}+ e^{-x}}$

$\sigma'(x) = - \frac{- e^{-x}}{(1+ e^{-x})^2} = \sigma(x)(1 - \sigma(x))$

$\tanh'(x) = \frac{(e^{x}+ e^{-x})(e^{x}+ e^{-x}) - (e^{x}- e^{-x})(e^{x}- e^{-x})}{(e^{x}+ e^{-x})2} = 1- \tanh2(x)$

they are used for efficient implementation of gradient

5.

Let gi be the gradient of i th point

gi = $(\omega . x_i - z_i)\, x_i$ if $| \omega . x_i - z_i | < 1$

$\mathrm{sgn}(\omega . x_i - z_i)\, x_i$ otherwise

then $\frac{\partial E}{\partial \omega} = \frac{1}{m}\sum_{i=1}^{m} g_i + \lambda\omega$

6.

During training, the progress of the performance, the magnitude of the gradient of performance and the number of validation checks are constantly updated. The magnitude of the gradient and the number of validation checks are used to terminate the training. The gradient will become very small as the training reaches a minimum of the performance. If the magnitude of the gradient is less than a specific value, the training will stop. The number of validation checks represents the number of successive iterations that the validation performance fails to decrease. If this number reaches the default value, the training will stop. The training will also stop if the performance function fails to decrease significantly over many iterations.