Introduction to Neural Networks : Important Equations

John A. Bullinaria - 2004

To study Neural Networks effectively, one has to deal with quite a lot of mathematics. The examination will not require you to carry out complex mathematical derivations, nor remember complex equations, but it may require you to understand and explain some of the following important equations from your lectures:

Basic neuron equation

$$out_j = f(\sum_{i=1}^{n} w_{ji} in_i - \theta_j)$$

Sigmoid/Logistic activation function

$$\text{Sigmoid}(x) = \frac{1}{1+e^{-x}}$$

Sum Squared Error Function

$$E_{sse}(\{w_{kl}\}) = \tfrac{1}{2} \sum_p \sum_j \left(targ_j^p - out_j(in_i^p) \right)^2$$

Gradient descent weight update equation

$$\Delta w_{kl} = -\eta \frac{\partial E(\{w_{ij}\})}{\partial w_{kl}}$$

Gradient descent for Single Layer Perceptron

$$\Delta w_{kl} = \eta \sum_p (targ_l - out_l).f'(\sum_n in_n w_{nl}).in_k$$

Back-propagation with momentum

$$delta_k^{(n)} = \left(\sum_k delta_k^{(n+1)}.w_{lk}^{(n+1)} \right).f'\left(\sum_j out_j^{(n-1)} w_{jk}^{(n)} \right)$$

$$\Delta w_{kl}^{(n)}(t) = \eta \sum_p delta_l^{(n)}(t).out_k^{(n-1)}(t) + \alpha.\Delta w_{kl}^{(n)}(t-1)$$

Regularization and weight decay

$$E_{reg} = E_{sse} + \lambda\Omega$$

$$\Omega = -\tfrac{1}{2}\sum_{j,i,m}(w_{ji}^{(m)})^2$$

Bias and Variance

$$\mathcal{E}_D\left[\left(\mathcal{E}[y \mid x_i] - net(x_i, W, D)\right)^2\right]$$

$$= \left(\mathcal{E}_D\left[net(x_i, W, D)\right] - \mathcal{E}[y \mid x_i]\right)^2 + \mathcal{E}_D\left[\left(net(x_i, W, D) - \mathcal{E}_D\left[net(x_i, W, D)\right]\right)^2\right]$$

$$= \quad (\text{bias})^2 \quad + \quad (\text{variance})$$

Radial Basis Functions

$$y_k(\mathbf{x}) = \sum_{j=0}^{M} w_{kj}\phi_j(\mathbf{x})$$

$$\phi_j(\mathbf{x}) = \exp\left(-\frac{\left\|\mathbf{x} - \boldsymbol{\mu}_j\right\|^2}{2\sigma_j{}^2}\right)$$

Mixtures of Experts

$$y(n) = \sum_{i=1}^{K} g_i(\mathbf{x}(n))y_i(n)$$

SOM Algorithm

$$d_j(\mathbf{x}) = \sum_{i=1}^{D}(x_i - w_{ji})^2$$

$$T_{j,I(\mathbf{x})}(t) = \exp(-S_{j,I(\mathbf{x})}^2 / 2\sigma^2(t))$$

$$\Delta w_{ji} = \eta(t)\ T_{j,I(\mathbf{x})}(t)\ (x_i - w_{ji})$$