COMP9444 Neural Networks and Deep Learning

Quiz 1 Answers (Perceptron Learning, Backpropagation and Variations)

1. Explain the difference between Perceptron Learning and Backpropagation.

   Perceptron Learning only applies to a Perceptron (one-layer neural network with step activation). Assume the function computed by the Perceptron is $g(w_0 + \Sigma_k w_k x_k)$ where $g()$ is the Heaviside step funtion, and η is the learning rate. If the output is 0 but should have been 1, η is added to the bias and ηx_k is added to each weight w_k. If the output is 1 but should have been 0, these values are instead subtracted rather than added.

   Backpropagation is a form of gradient descent, which can be applied to multi-layer neural networks provided the activation function is (mostly) differentiable. The derivative $\partial E/\partial w$ of the cost function E with respect to each weight w is calculated, and $\eta\ \partial E/\partial w$ is subtracted from w.

2. When training a neural network by backpropagation, what happens if the learning rate is too low? What happens if it is too high?

   If the learning rate is too low, the training will be very slow. If it is too high, the training may become unstable and fail to learn the task successfully.

3. Explain why rescaling of inputs is sometimes necessary for neural networks.

   The differential of each weight in the first layer gets multiplied by the value of its corresponding input. Therefore, the network may give undue emphasis to inputs of larger magnitude. Rescaling encourages all inputs to be treated with equal importance.

4. Explain what is meant by "overfitting" in neural networks, and list three different methods for avoiding it.

   Overfitting is where the training set error continues to reduce, but the test set error stalls or increases. This can be avoided by
   a. reducing the number of neurons or connections in the network
   b. early stopping, with a validation set
   c. weight decay (discussed in Week 3 lecture)

5. What is the difference between batch learning and online learning?

   For batch learning, the differentials accruing from each training item are aggregated, and the weights are only updated at the end of the batch. The entire training set might form a single batch, or it may be divided into a number of subsets called "minibatches". For online learning (also called stochastic gradient descent) the weights are updated after each training item, using the differentials generated by that item.

6. Write the formula for a Gaussian distribution with mean μ and standard deviation σ.

   $P(x) = \exp(-(x-\mu)^2/2\sigma^2)/(\text{sqrt}(2\pi)\sigma)$