

COMP9444 Neural Networks and Deep Learning**Quiz 2 (Variations on Backprop)**

This is an optional quiz to test your understanding of the material from Week 3.

1. Write the formulas for these three error functions: squared error, cross entropy, softmax (remember to define any variables you use).

Assume z is the actual output and t is the target output.

squared error: $E = (z-t)^2/2$

cross entropy: $E = -t \log(z) - (1-t) \log(1-z)$

softmax: $E = -(z_i - \log \sum_j \exp(z_j))$, where i is the correct class.

2. Write the formula for Bayes' Rule, in terms of a cause A and an effect B.

$$P(A|B) = P(B|A)P(A)/P(B)$$

3. In the context of supervised learning, explain the difference between Maximum Likelihood estimation and Bayesian inference.

In Maximum Likelihood estimation, the hypothesis $h \in H$ is chosen which maximizes the conditional probability $P(D | h)$ of the observed data D , conditioned on h . In Bayesian inference, the hypothesis $h \in H$ is chosen which maximizes $P(D | h)P(h)$, where $P(h)$ is the prior probability of h .

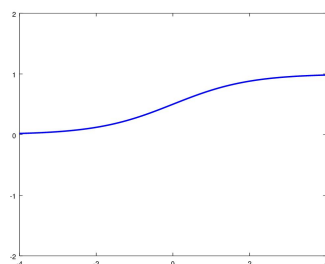
4. Briefly explain the concept of momentum, as an enhancement for gradient descent.

A running average of the differentials for each weight is maintained and used to update the weights as follows:

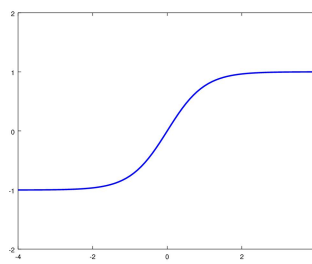
$$\delta w = \alpha \delta w + (1-\alpha) dE/dw, w = w - \eta \delta w$$

The constant α with $0 \leq \alpha < 1$ is called the momentum.

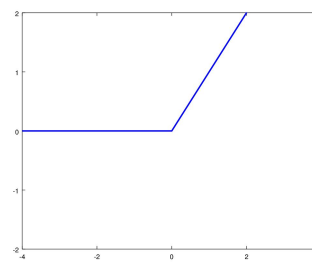
5. Sketch the following activation functions, and write their formula: sigmoid, tanh, ReLU.



$$y = 1/(1 + \exp(-x))$$



$$y = \tanh(x)$$



$$y = 0, \text{ if } x \leq 0$$

$$y = x, \text{ if } x > 0$$