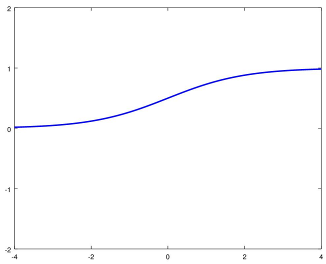
COMP9444 Neural Networks and Deep Learning

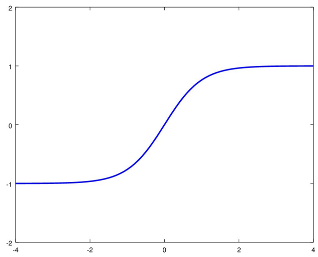
Quiz 3 (Convolutional Networks)

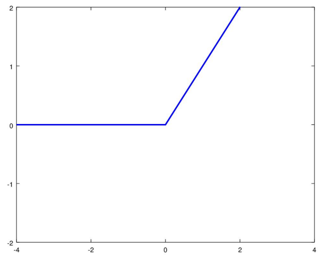
This is an optional quiz to test your understanding of the material from Weeks 3 and 4.

---

1. Sketch the following activation functions, and write their formula: Sigmoid, Tanh, ReLU.







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

$y = \tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$

$y = 0$, if $x \le 0$
$y = x$, if $x > 0$

2. Explain how Dropout is used for neural networks, in both the training and testing phase.

During each minibatch of training, a fixed percentage (usually, one half) of nodes are chosen to be inactive. In the testing phase, all nodes are active but the activation of each node is multiplied by the same percentage that was used in training.

3. Explain what is meant by Overfitting in neural networks, and list four 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
   d. dropout

4. Write the formula for the Softmax loss function

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

5. Write the formula for activation $Z^i_{j,k}$ of the node at location (j,k) in the i^{th} filter of a Convolutional neural network which is connected by weights $K^i_{l,m,n}$ to all nodes in an $M \times N$ window from the L channels in the previous layer, assuming bias weights are included and the activation function is $g()$. How many free parameters would there be in this layer?

$$Z^i_{j,k} = g(b^i + \Sigma_l \Sigma_{0 \le m < M} \Sigma_{0 \le n < N} K^i_{l,m,n} V^l_{j+m,k+n})$$

The number of free parameters is: F × (1 + L × M × N) where F is the number of filters in this layer.

6. If the previous layer has size $J \times K$, and a filter of size $M \times N$ is applied with stride s and zero-padding of width P, what will be the size of the resulting convolutional layer?

   $(1 + (J+2P-M)/s) \times (1 + (K+2P-N)/s)$

7. If max pooling with filter size F and stride s is applied to a layer of size $J \times K$, what will be the size of the resulting (downsampled) layer?

   $(1 + (J-F)/s) \times (1 + (K-F)/s)$
