Chapter 22: Learning by training neural networks

*Italics denotes where I have paraphrased and it is not same wording in book*

Each link is associated with a weight which determines the nature and strength of one nodes influence on another. One nodes influence on another is the product of the influencing neurons output value time the connecting links weight. Large positive weight = strong excitation.

Each node combines separate influences received on its input links into an overall influence using an activation function. Simple activation function example is just using a threshold function, output is either 0 or 1 based on if above or below threshold.

# Back propagation does hill climbing by gradient ascent

Adjust weights to improve output of NN – move in direction of most rapid performance improvement by varying all weights simultaneously *in proportion to the improvement observed.* This is gradient ascent. BP computes changes to weights in final layer first then reuses much of the computation to *change weights in preceding layers, sequentially.*