Perceptron model theory

Max Cotton

1 Setup

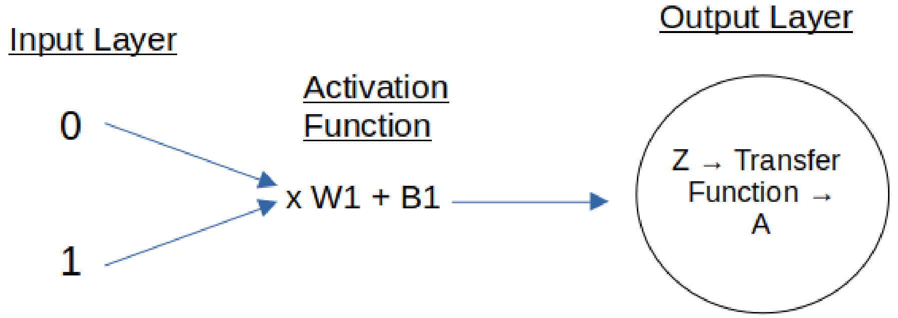

Figure 1: This model uses no hidden layers and is known as a Perceptron Artificial Neural Netwok.

- Where W is the array of weights, initially set to zeroes, and B is the array of biases, initially set to zeros
- Z is the dot product of the W array and the input array (X) summed with the B array
- A is sigmoid(Z), which is the prediction
  - Where $sigmoid(Z) = \frac{1}{1+e^{-Z}}$

2 Forward Propagation

For each epoch the input array of values, is fed through the network to obtain a prediction A

3 Back Propagation

- Once a prediction is obtained, you then move back through the network adjusting the weights and the bias
- The "Loss" (L) (how wrong the prediction is) can be calculated with the Loss function, which calculates the average difference between the prediction and the actual value (Y), the average is found by summing the

result for each prediction then dividing by the number of predictions (m). This shows how well the network is performing.

  - Where $L = -(\frac{1}{m}) * \sum(Y * log(A) + (1-Y) * log(1-A))$ and "log()" is the natural logarithm (e is the base)

- The weights and the bias are adjusted via Gradient Descent, which aims to get the minimum Loss value, with the following formulae:
  - $W = W - learningRate * \frac{\partial L}{\partial W}$
    * Where $\frac{\partial L}{\partial W} = X * (\frac{1}{m})(A - Y)$
  - $B = B - learningRate * \frac{\partial L}{\partial B}$
    * Where $\frac{\partial L}{\partial B} = (\frac{1}{m})(A - Y)$

4 Derivations for $\frac{\partial L}{\partial W}$ and $\frac{\partial L}{\partial B}$

- Functions used so far:
  1. $Z = W.X + B$
  2. $A = \frac{1}{1+e^{-Z}}$
  3. $L = -(\frac{1}{m}) * \sum(Y * log(A) + (1-Y) * log(1-A))$
- $\frac{\partial L}{\partial W}$ derivation:
  - By the Chain rule, $\frac{\partial L}{\partial W} = \frac{\partial L}{\partial A} * \frac{\partial A}{\partial Z} * \frac{\partial Z}{\partial W}$
  - Using function 3, $\frac{\partial L}{\partial A} = (-\frac{1}{m})(\frac{Y-A}{A*(1-A)})$
  - Using function 2, $\frac{\partial A}{\partial Z} = A * (1-A)$
  - Using function 1, $\frac{\partial Z}{\partial W} = X$
  - =¿ $\frac{\partial L}{\partial W} = (-\frac{1}{m})(\frac{Y-A}{A*(1-A)}) * A * (1-A) * X$
    $= X * (\frac{1}{m})(A - Y)$
- $\frac{\partial L}{\partial B}$ derivation:
  - By the Chain rule, $\frac{\partial L}{\partial B} = \frac{\partial L}{\partial A} * \frac{\partial A}{\partial Z} * \frac{\partial Z}{\partial B}$
  - Derived above, $\frac{\partial L}{\partial A} = (-\frac{1}{m})(\frac{Y-A}{A*(1-A)})$
  - Derived above, $\frac{\partial A}{\partial Z} = A * (1-A)$
  - Using function 1, $\frac{\partial Z}{\partial B} = 1$
  - =¿ $\frac{\partial L}{\partial W} = (-\frac{1}{m})(\frac{Y-A}{A*(1-A)}) * A * (1-A)$
    $= (\frac{1}{m})(A - Y)$