Notes on Machine learning

Matthew

today

1 Basics - Neural Networks

Cost function C:

$$C(w,b) = \frac{1}{2n}\sum_x ||y(x) - a||^2$$

Change in C is modelled approximately by:

$$\Delta C \approx \frac{\delta C}{\delta v_1}\Delta v_1 + \frac{\delta C}{\delta v_2}\Delta v_2$$

The gradient vector of C is equivalent to:

$$\nabla C \equiv \left(\frac{\delta C}{\delta v_1}, \frac{\delta C}{\delta v_2}\right)^T$$

Therefore a change in C can be approximated by $\Delta C \approx \nabla C \cdot \Delta v$. As we are trying to minimise the cost value of C, we can fix Δv to a value that ensures ΔC will always be negative:

$$\Delta v = -\eta \nabla C$$

Where η is a small positive value known as the learning rate.
This means that:

$$\begin{aligned} \Delta C &= \nabla C \cdot \Delta v \\ &= \nabla C \cdot -\eta \nabla C \\ &= -\eta ||\nabla C||^2 \end{aligned}$$

i.e. the change in C can always be negative, towards the local minima. The speed of gradient descent is now:

$$v \to v' = v - \eta \nabla C$$