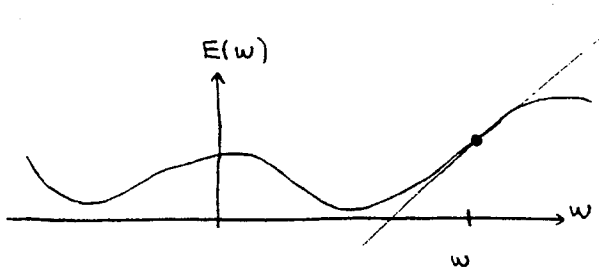
Apr 1990 Gradient Descent – Mathematical View

Neil E Cotter

$$\Delta w = -\eta \frac{\partial E(w)}{\partial w}$$

or $$\Delta \vec{w} = -\eta \nabla E(\vec{w})$$



positive slope ⇒ negative Δw will give lower E.

Observe that $\frac{\partial E}{\partial w} \approx \frac{\Delta E}{\Delta w}$ = change in E for a change in w

If $\frac{\Delta E}{\Delta w} > 0$ then positive change in w implies positive change in ΔE.

To lower E, we should therefore use Δw negative, which is what $-\eta \frac{\partial E(w)}{\partial w}$ does. $(\eta > 0)$

If $\frac{\Delta E}{\Delta w} < 0$ then positive change in w implies negative change in ΔE.

To lower E, we should therefore use Δw positive, which is what $-\eta \frac{\partial E}{\partial w}$ does.

Conclusion: Gradient descent always moves us toward lower E, (if we make small steps, $\eta \ll 1$).

If $\frac{\partial E(w)}{\partial w}$ is large we stand to lower E more by making a larger step in w.