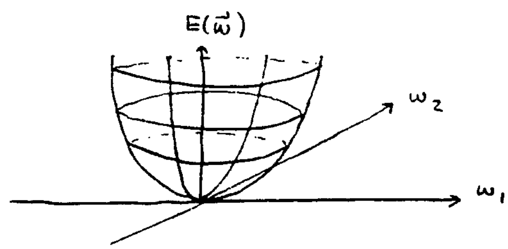
19 Apr 1990 Gradient Descent – Quadratic Bowls

Neil E Cotter

claim: Near a minimum, every infinitely differentiable function of n variables is approximately a quadratic bowl.



Quadratic bowl for 2 variables. Bottom of bowl need not be at height = 0. Need not be at $w_1 = w_2 = 0$, either

pf: Let $\vec{w}_0$ be at a minimum of $E(\vec{w})$

Consider a small perturbation of $\vec{w}_0$:

$\vec{w} = \vec{w}_0 + \vec{\epsilon}$ where $\vec{\epsilon}$ is small vector

Then by Taylor series:

$$E(\vec{w}) = E(\vec{w}_0 + \vec{\epsilon}) = E(\vec{w}_0) + \nabla_{\vec{w}} E(\vec{w})\Big|_{\vec{w}=\vec{w}_0} \circ \vec{\epsilon} + \frac{\vec{\epsilon}^T H \vec{\epsilon}}{2} + O(|\vec{\epsilon}|^3)$$

since $O(|\epsilon|^3)$ small " $\approx E(\vec{w}_0) + \nabla_{\vec{w}} E(\vec{w})\Big|_{\vec{w}=\vec{w}_0} \circ \vec{\epsilon} + \frac{\vec{\epsilon}^T H \vec{\epsilon}}{2}$

But $\nabla_{\vec{w}} E(\vec{w})\Big|_{\vec{w}=\vec{w}_0} = 0$ since $E(\vec{w})$ is flat at bottom of bowl.

So we have $\underbrace{E(\vec{w}_0 + \vec{\epsilon})}_{} = \underbrace{E(\vec{w}_0)}_{\text{const}} + \underbrace{\vec{\epsilon}^T H \vec{\epsilon}/2}_{\text{quadratic}}$

i.e. we have quadratic bowl.

Note: H is Hessian (matrix) $\equiv \left[\frac{\partial^2 E(\vec{w})}{\partial w_j \partial w_i}\right]\Bigg|_{\vec{w}=\vec{w}_0}$