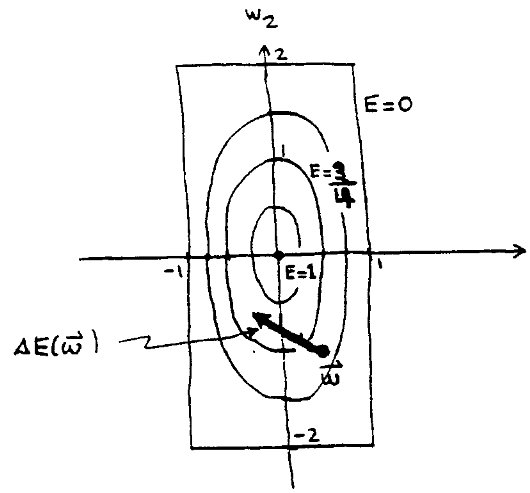
Apr 1990 Gradient Descent — Calculating Gradients

Neil E Cotter

ex: Find $\nabla E(w_1, w_2)$ where $E(w_1,w_2) = (1-w_1^2)(1-\frac{w_2^2}{4})$

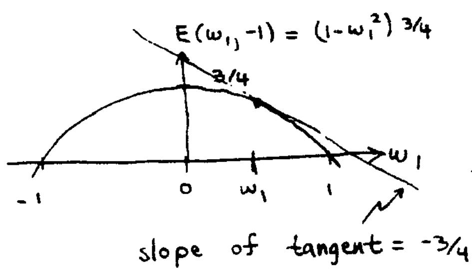


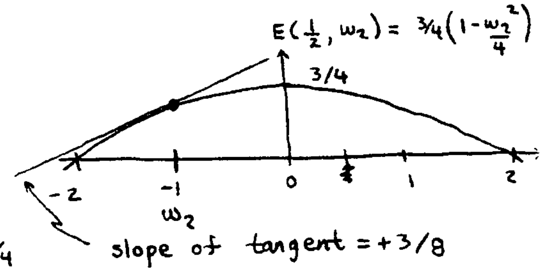
$$\nabla E(w_1,w_2) \equiv \begin{bmatrix} \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial w_1} (1-w_1^2)(1-\frac{w_2^2}{4}) \\ \frac{\partial}{\partial w_2} (1-w_1^2)(1-\frac{w_2^2}{4}) \end{bmatrix} = \begin{bmatrix} -2w_1(1-\frac{w_2^2}{4}) \\ -\frac{2w_2}{4}(1-w_1^2) \end{bmatrix}$$

Consider $\vec{w} \equiv (w_1, w_2) = (\frac{1}{2}, -1)$

Then $\nabla E(\frac{1}{2}, -1) = \begin{bmatrix} -2(\frac{1}{2})(1-\frac{(-1)^2}{4}) \\ -\frac{2(-1)}{4}(1-(\frac{1}{2})^2) \end{bmatrix} = \begin{bmatrix} -3/4 \\ 3/8 \end{bmatrix}$ slope in w_1 dir / slope in w_2 d

Estimate $\partial E/\partial w_1$, $\partial E/\partial w_2$ from func profiles as a check:



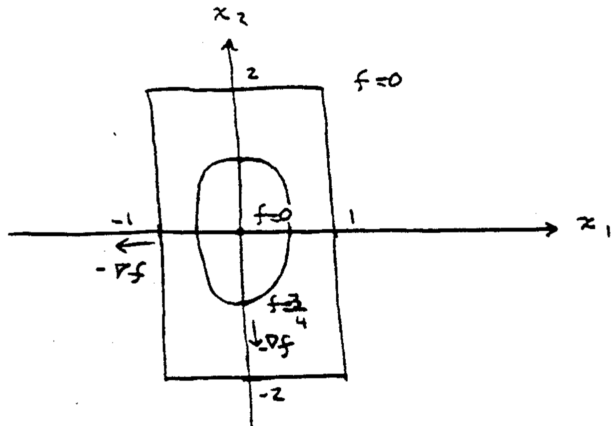


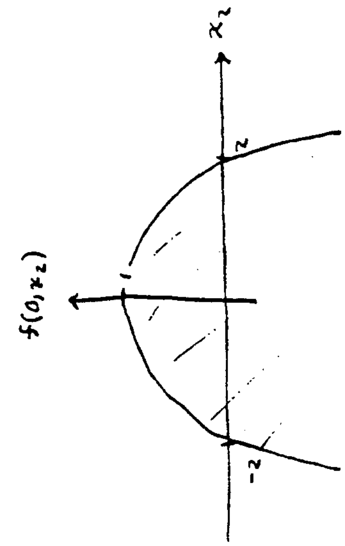
Crude sketch of profiles suggests $\nabla E = \begin{bmatrix} -3/4 \\ 3/8 \end{bmatrix}$ is plausib

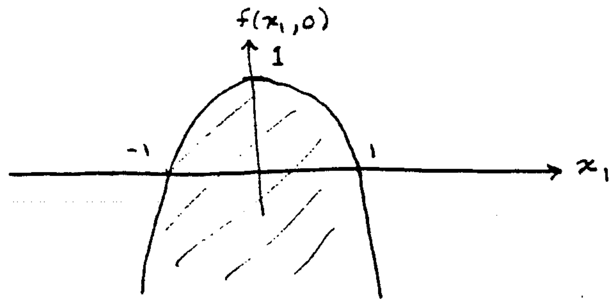
May 1988 Gradient Descent – Calculating Gradients (cont.)

Neil E Cotter *ex:* Calc gradient

$$f(x_1, x_2) = \left[1 - x_1^2\right]\left[1 - \frac{x_2^2}{4}\right]$$







- *Cross sections are parabolas*

$$\nabla f = \begin{pmatrix} \dfrac{\partial}{\partial x_1}\left[1 - x_1^2\right]\left[1 - \dfrac{x_2^2}{4}\right] \\[2ex] \dfrac{\partial}{\partial x_2}\left[1 - x_1^2\right]\left[1 - \dfrac{x_2^2}{4}\right] \end{pmatrix} = \begin{pmatrix} -2x_1\left[1 - \dfrac{x_2^2}{4}\right] \\[2ex] \left[1 - x_1^2\right]\left(-2\dfrac{x_2}{4}\right) \end{pmatrix}$$

$$\text{at } (-1,0) \quad -\nabla f = -\begin{pmatrix} -2(-1)\left[1 - 0^2/4\right] \\[2ex] \left[1 - -1\right] -2\cdot 0/4 \end{pmatrix} = -\begin{pmatrix} 2 \\[2ex] 0 \end{pmatrix} = \begin{pmatrix} -2 \\[2ex] 0 \end{pmatrix}$$

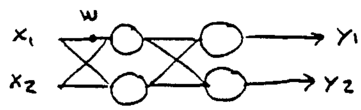
$$\text{at } (0,-1) \quad -\nabla f = -\begin{pmatrix} -2\cdot 0\left[1 - \dfrac{(-1)^2}{4}\right] \\[2ex] \left[1 - 0^2\right]\dfrac{(-2)(-1)}{4} \end{pmatrix} = -\begin{pmatrix} 0 \\[2ex] 1/4 \end{pmatrix} = \begin{pmatrix} 0 \\[2ex] -1/4 \end{pmatrix}$$

- *Gentler slope for gradient at* $(0,-1)$ *as can be seen in the picture.*

Apr 1990 Gradient Descent — Calculating Gradients (cont.)

Neil E Cotter

I. Multiple Output Net



t_1 target value
t_2 " "

Define error for multiple outputs:

$$E \equiv \frac{1}{2} \sum_{k=1}^{n} (t_k - y_k)^2 \quad \text{for } n \text{ outputs}$$

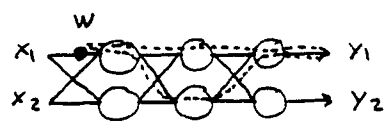
ex: $n=2$ $$E = \frac{1}{2}(t_1 - y_1)^2 + \frac{1}{2}(t_2 - y_2)^2$$

A synaptic weight, w, can affect multiple outputs:

$$\frac{\partial E}{\partial w} = \frac{\partial}{\partial w} \frac{1}{2} \sum_{k=1}^{n} (t_k - y_k)^2 = \sum_{k=1}^{n} (t_k - y_k) \frac{\partial y_k}{\partial w}$$

$\therefore$ we calculate affect on each output and then sum.

II. Multiple Pathways to Output



t_1 target val

Dotted paths show how w affects y_1 in two ways.

We can write a symbolic formula for y_1:

$$y_1 = f(g_1(w), g_2(w))$$

Then from multivariable calculus we have:

$$\frac{\partial y_1}{\partial w} = \frac{\partial}{\partial w} f(g_1(w), g_2(w)) = \frac{\partial f}{\partial g_1} \frac{\partial g_1(w)}{\partial w} + \frac{\partial f}{\partial g_2} \frac{\partial g_2(w)}{\partial w}$$

In general we have $$\frac{\partial y}{\partial w} = \frac{\partial}{\partial w} f(g_1(w), \ldots, g_n(w)) = \sum_{k=1}^{n} \frac{\partial f}{\partial g_k} \frac{\partial g_k}{\partial w}$$

Neil E Cotter

ex: Given: $y(g_1(w), g_2(w)) = g_1(w) \cdot g_2(w)$

and $g_1(w) = w^2 \quad g_2(w) = w^3$

Find: $\dfrac{\partial y}{\partial w}$ Use multiple pathways rule (II)

sol'n: $$\frac{\partial y}{\partial w} = \sum_{k=1}^{2} \frac{\partial y}{\partial g_k} \frac{\partial g_k}{\partial w}$$

$$\frac{\partial y}{\partial g_1} = \frac{\partial}{\partial g_1} g_1 \cdot g_2 = g_2$$

$$\frac{\partial y}{\partial g_2} = \frac{\partial}{\partial g_2} g_1 g_2 = g_1$$

(~~treat~~ funcs g_1, g_2 like variables; g_2 acts like const. multiplying g_1, and g_1 acts like a variable such as x in $\frac{\partial f(x)}{\partial x}$)

Now we re-introduce w and note that g_2 is really $g_2(w)$, i.e. g_2 evaluated at point w.

$$\frac{\partial g_1}{\partial w} = \frac{\partial}{\partial w} w^2 = 2w$$

$$\frac{\partial g_2}{\partial w} = \frac{\partial}{\partial w} w^3 = 3w^2$$

So $$\frac{\partial y}{\partial w} = \underset{\frac{\partial y}{\partial g_1}}{g_2(w)}\ \underset{\frac{\partial g_1}{\partial w}}{2w} + \underset{\frac{\partial y}{\partial g_2}}{g_1(w)}\ \underset{\frac{\partial g_2}{\partial w}}{3w^2}$$

$$= w^3 \cdot 2w + w^2 \cdot 3w^2$$

$$= 5w^4$$

Can check this: $y = g_1 g_2 = w^2 w^3 = w^5$, $\dfrac{\partial y}{\partial w} = 5w^4$ ✓

Not always convenient to find $\partial y/\partial w$ this way.