Derivative of a scalar function $f(w)\ ; w \in \mathbb{R}$

- $\frac{df}{dw}$ = How the functions changes at w as w changes
- slope of tangent at w

If f is a function of more than one variable, $f(w_1,\ w_2,\ w_3)$ $[f(\mathbf{w})]$

- $\left.\frac{\partial f}{\partial w_1}\right|_{(a,\,b,\,c)}$ = How the function changes at $(a,\ b, c)$ if we move in the direction of w_1

Gradient of a scalar function $f(w_1,\ w_2,\ w_3)$ $[f(\mathbf{w})]$

$$\left.\nabla f\right|_{(a,\,b,\,c)} = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \frac{\partial f}{\partial w_3} \end{bmatrix}$$

- It is the direction in which function change is maximum at $(a,\ b,\ c)$
- What is $|\nabla f|$??

NOTE: We will be following numerator layout notations

Derivative of a scalar valued function $f(w_1,\ w_2,\ w_3)$ w.r.t. a vector $\mathbf{w} = [w_1\ \ w_2\ \ w_w]^T$

$$\frac{df}{d\mathbf{w}} = \left.\left[\frac{\partial f}{\partial w}, \frac{\partial f}{\partial w_2}, \frac{\partial f}{\partial w_3}\right]\right|_{(a,\,b,\,c)} = (\nabla f)^T$$

- What is directional derivatives??

Scalar by matrix

Let $\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1d} \\ w_{21} & w_{22} & \dots & w_{2d} \\ \vdots & \vdots & & \vdots \\ w_{d1} & w_{d2} & \dots & w_{dd} \end{bmatrix}$ and y be a scalar valued function of $w's$

- $$\frac{\partial y}{\partial \mathbf{W}} = \begin{bmatrix} \frac{\partial y}{\partial w_{11}} & \frac{\partial y}{\partial w_{21}} & \dots & \frac{\partial y}{\partial w_{d1}} \\ \frac{\partial y}{\partial w_{12}} & \frac{\partial y}{\partial w_{22}} & \dots & \frac{\partial y}{\partial w_{d2}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y}{\partial w_{1d}} & \frac{\partial y}{\partial w_{2d}} & \dots & \frac{\partial y}{\partial w_{dd}} \end{bmatrix}$$

Vector by scalar

Let $\mathbf{y} = [y_1 \;\; y_2 \;\; y_3 .... y_n]^T$

- $\frac{d\mathbf{y}}{dw} = \begin{bmatrix} \frac{dy_1}{dw} & \frac{dy_2}{dw} & \frac{dy_3}{dw} & \dots & \frac{dy_n}{dw} \end{bmatrix}^T$
- Called a tangent vector

Vector by vector

Let $\mathbf{y} = [y_1 \;\; y_2 \;\; y_3 .... y_n]^T$ and $\mathbf{w} = [w_1 \;\; w_2 \;\; w_w \;\; \dots \;\; w_d]^T$

- A $n \times d$ matrix called Jacobian matrix
- $$\frac{d\mathbf{y}}{d\mathbf{w}} = \begin{bmatrix} \frac{dy_1}{dw_1} & \frac{dy_1}{dw_2} & \dots & \frac{dy_1}{dw_d} \\ \frac{dy_2}{dw_1} & \frac{dy_2}{dw_2} & .. & \frac{dy_2}{dw_d} \\ \vdots & \vdots & & \vdots \\ \frac{dy_n}{dw_1} & \frac{dy_n}{dw_2} & .. & \frac{dy_n}{dw_d} \end{bmatrix}$$

-

$$\frac{\partial \mathbf{w}}{d\mathbf{w}} = \mathbf{I}_d$$

$$\frac{\partial \mathbf{w}^T}{\partial \mathbf{w}} = [1,\ 0,\ldots 0 \quad 0,\ 1,\ 0,\ ..0 \quad 0,\ 0,\ \ldots,\ 1]]$$

Some Identities:

$$\frac{d(\mathbf{A}\mathbf{w})}{d\mathbf{w}} = A$$

$$\frac{d\left(\mathbf{w}^{\mathrm{T}}\mathbf{A}\right)}{d\mathbf{w}} = A^T$$

$$\frac{d\left(\mathbf{w}^T\mathbf{A}\mathbf{w}\right)}{d\mathbf{w}} = \mathbf{w}^T\left(\mathbf{A}^T + \mathbf{A}\right)$$

Loss function

$$\begin{aligned} L(\mathbf{w}) &= \left(\mathbf{X}^T\mathbf{w} - \mathbf{y}\right)^T\left(\mathbf{X}^T\mathbf{w} - \mathbf{y}\right) \\ &= \left(\mathbf{w}^{\mathrm{T}}\mathbf{X} - \mathbf{y}^T\right)\left(\mathbf{X}^T\mathbf{w} - \mathbf{y}\right) \\ &= \mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w} - \mathbf{y}^T\mathbf{X}^T\mathbf{w} - \mathbf{w}^T\mathbf{X}\mathbf{y} + \mathbf{y}^T\mathbf{y} \\ &= \mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w} - 2\,\mathbf{y}^T\mathbf{X}^T\mathbf{w} + \mathbf{y}^T\mathbf{y} \end{aligned}$$

$$\begin{aligned} \nabla L &= \left(\frac{\partial L}{\partial \mathbf{w}}\right)^T = \left(2\mathbf{w}^{\mathrm{T}}\mathbf{X}\mathbf{X}^T - 2\,\mathbf{y}^T\mathbf{X}^T\right)^T \\ &= 2\mathbf{X}\mathbf{X}^T\mathbf{w} - 2\mathbf{X}\mathbf{y} \end{aligned}$$