[Theory and Practice of Deep Learning] Theory HW #02

Nigel Chan, 1002027

Task 1 : Disastrous Derivativez

Compute directional derivative $Df(X)[H]$ in direction H :

a) $f(X) = Xa$, $Df(X)[H] = Ha$, $H \in \mathbb{R}^{d \times k}$

b) $f(X) = XX^T$, $Df(X)[H] = HX^T + X^T H$, $H \in \mathbb{R}^{d \times n}$

c) $f(X) = XCX$, $Df(X)[H] = HCX + XCH$

d) $f(X) = CXBX^TAX$, $Df(X)[H] = CHBX^TAX + CXBH^TAX + C\overset{X}{\cancel{H}}BX^TAH$

e) Qns: Can something similar be done with a linear, a bilinear or a trilinear function here?

Ans: Yes.

f) $f(X) = \begin{pmatrix} 1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & x_2^3 \\ \sin x_2 & x_1 \end{pmatrix}$

$= \begin{pmatrix} 1 + x_2 \sin x_2 & x_2^3 + x_1 x_2 \end{pmatrix}$

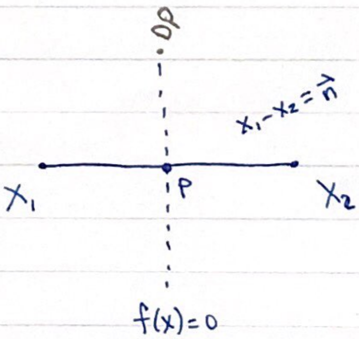
$\nabla f(X) = \begin{pmatrix} 0 & x_2 \\ \sin x_2 + x_2 \cos x_2 & 3x_2^2 + x_1 \end{pmatrix}$

Following the formula : $\nabla_{\vec{v}} f(x) = \nabla f(x) \cdot \vec{v}$

$\therefore Df(X)[H] = H^T \begin{pmatrix} 0 & x_2 \\ \sin x_2 + x_2 \cos x_2 & 3x_2^2 + x_1 \end{pmatrix}$

Task 2 : n-dim Hyperplanes are a piece of bunny

Part 2.1, 3dims



$X_1 = \begin{bmatrix} -5 \\ 1 \\ 2 \end{bmatrix}$, $Y_1 = +1$, $X_2 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$, $Y_2 = -1$

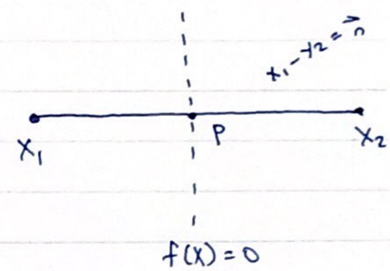
the normal $\vec{n}$ is the vector from X_2 to X_1,

$\vec{n} = X_1 - X_2$

$= \begin{bmatrix} -6 \\ -1 \\ 5 \end{bmatrix}$

let $\vec{p}$ be the point on the hyperplane existing between x_1 and x_2,

$$\vec{p} = \frac{x_1 + x_2}{2} = \begin{bmatrix} -2 \\ 1.5 \\ -0.5 \end{bmatrix}$$



let $ax + by + cz = D$ be the hyperplane,

sub $\vec{n}$ and $\vec{p}$ values in,

$$(-6)(-2) + (-1)(1.5) + (5)(-0.5) = D$$

$$D = 8$$

By definition, $f(x) = 0$,

Hence, ~~$-6x - y + 5z$~~

$$f(x) = -6x - y + 5z - 8$$

$$= \begin{bmatrix} -6 \\ -1 \\ 5 \end{bmatrix} \cdot \vec{x} - 8$$

$\therefore \vec{w} = \begin{bmatrix} -6 \\ -1 \\ 5 \end{bmatrix}$, $b = -8$

part 2.2, 5 dims I

normal $\vec{n}$ is vector from x_2 to x_1,

$$\vec{n} = x_1 - x_2 = \begin{bmatrix} -2 \\ -1 \\ 4 \\ -7 \\ 5 \end{bmatrix}$$

let $\vec{p}$ be point on hyperplane,

$$\vec{p} = \frac{x_1 + x_2}{2} = \begin{bmatrix} 2 \\ 0.5 \\ -1 \\ 1.5 \\ 3.5 \end{bmatrix}$$

let $ax_1 + bx_2 + cx_3 + dx_4 + ex_5 = B$ be the hyperplane

sub $\vec{n}$ and $\vec{p}$ values,

$$\begin{bmatrix} -2 \\ -1 \\ 4 \\ -7 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0.5 \\ -1 \\ 1.5 \\ 3.5 \end{bmatrix} \quad -4 - 0.5 - 4 - 10.5 + 17.5 = B$$

$$B = -1.5$$

By definition, $f(x) = 0$

$$f(x) = -2x_1 - x_2 + 4x_3 - 7x_4 + 5x_5 + 1.5$$

$$= \begin{bmatrix} -2 \\ -1 \\ 4 \\ -7 \\ 5 \end{bmatrix} \cdot \vec{x} + 1.5$$

$\therefore \vec{w} = \begin{bmatrix} -2 \\ -1 \\ 4 \\ -7 \\ 5 \end{bmatrix}$, $b = 1.5$

part 2.3, 5 dims II

I. Choose $\vec{w} = \begin{bmatrix} 6 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$

II. To check that w is not parallel to $x_2 - x_3$, ~~we use~~

let $\vec{z} = x_2 - x_3 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 1 \\ -1 \end{bmatrix}$

if $\cos^{-1} \frac{w \cdot z}{\|w\| \|z\|} = \theta \;!= 0$, w is not parallel.

sub in points,

$\vec{w} \cdot \vec{z} = 1$, $\|w\| = \sqrt{66}$, $\|z\| = 4$

$\cos^{-1} \frac{1}{4\sqrt{66}} = 88.2 \;!= 0$

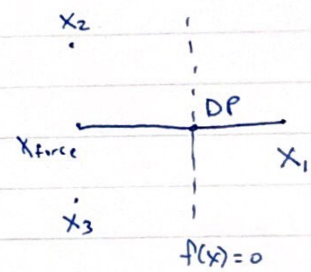
$\therefore w$ is not parallel to $x_2 - x_3$

To check that w is not orthogonal to $x_1 - x_3$, we check for $w \cdot (x_1 - x_3) \;!= 0$

$\vec{w} \cdot (\vec{x_1} - \vec{x_3})$

$= \begin{bmatrix} 6 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$

$= 7 \;!= 0$, Hence w is not orthogonal to $x_1 - x_3$.



III. $\vec{w_2} = w - (w \cdot z) \frac{z}{\|z\|^2}$

$= \begin{bmatrix} 6 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 1/16 \\ 1/8 \\ -3/16 \\ 1/16 \\ -1/16 \end{bmatrix}$

$= \begin{bmatrix} 95 \\ 14 \\ 35 \\ 47 \\ 65 \end{bmatrix}$

IV. $\vec{w_2} \cdot \vec{z} = \begin{bmatrix} 95 \\ 14 \\ 35 \\ 47 \\ 65 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ -3 \\ 1 \\ -1 \end{bmatrix}$

$= 0$ $\therefore w_2$ is orthogonal to $x_2 - x_3 = z$

V. let X_{force} be the point between x_2 and x_3

$X_{force} = \frac{x_2 + x_3}{2} = \begin{bmatrix} 2.5 \\ 0 \\ -1.5 \\ 4.5 \\ 1.5 \end{bmatrix}$

let DP be the point on the hyperplane between X_1 and X_{force}

$DP = \frac{X_{force} + X_1}{2} = \begin{bmatrix} 1.75 \\ 0 \\ -0.25 \\ 5.25 \\ 2.25 \end{bmatrix}$

let $\vec{p} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$

$(\vec{p} - \vec{DP}) \cdot W_2 = 0$

$95x_1 + 14x_2 + 35x_3 + 47x_4 + 65x_5 - 550.5 = 0$

$f(x) = 0$

$= \begin{bmatrix} 95 \\ 14 \\ 35 \\ 47 \\ 65 \end{bmatrix} \vec{p} - 550.5$ $\therefore$ bias $= -550.5$

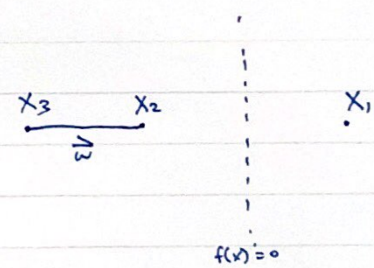


Fig. A

VI If for example, there is the case where the points x_1, x_2, x_3 are in line (look at fig A), w, being parallel to $x_2 - x_3$, will become unsuitable as a separating hyperplane