1) $\min_{h \in H} \sum_{i=1}^{N} [\text{Sign}(h(x)) \neq y_i]$ where $[\cdot]$ = Indicator Function

ⓐ $\max\{0, 1 - y_i h(x_i)\}$

y	$h(x)$	Loss
+1	high	Low
+1	low	high
−1	high	high
−1	Low	low

Thus this is a good proposal

ⓑ $\min\{0, 1 - y_i h(x_i)\}$

This is not a good loss function

ⓒ $$\frac{\exp(-y_i h(x_i))}{1 + \exp(-y_i h(x_i))}$$

$$= \frac{1}{1 + \exp(y_i h(x_i))}$$

y_i	$h(x_i)$	loss
+1	high	low
+1	low	high
−1	high	high
−1	Low	low

Thus ⓒ is also a good proposal

ⓓ It is easy to see that loss in qn ⓓ = 1 - (loss in qn ⓒ) and so answer is NO.

② any of the losses in {1a, 1b, 1c} would do

③ only ⓐ and ⓑ

④ ⓐ we guess the majority label in Training data

$\hat{c} = +1$

ⓑ since Error is calculated on the test data

$\text{Error}(\hat{h}) = \text{Error}(h^*) = 0.5$

Answer $= 0$

⑤ Because test set has 60% +ve samples any scheme that assigns high value to +ve labelled examples in Training data works

⑥ ⓐ yes, it is valid Loss because

$w^T x$	y_i	loss
high	+1	low
high	0	high
low	+1	high
low	0	low

⑦ when loss = 0 $h(x) = \{+1, -1\}$ which is only possible when $\|w\| = \infty$

⑧ If $h(x) = \frac{1}{1 + \exp(+w^T x)}$

we can change the Loss function as

$$\sum_i - \{ y_i \log(1 - h(x_i)) + (1 - y_i) \log h(x_i) \}$$

⑨ ⓐ If $\tau = 0$ the model always predicts +1

ⓑ If $\tau = 1$ the model always predicts 0

ⓒ If $\tau = 0.5$ the model predicts a mix of $\{+1, 0\}$. But the test accuracy is not optimal

ⓓ $$\tau^* = \underset{\tau \in [0,1]}{\operatorname{argmin}} \sum_{(x_i, y_i) \in D_{Test}} \left[\operatorname{Sign}(h(x_i) - \tau) \neq y_i \right]$$

8 Linearity of functions

(a) Linear in x and Linear in ω

$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\bar{\omega} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

$\alpha\bar{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix}$ $\alpha\bar{\omega} = \begin{bmatrix} \alpha w_1 \\ \alpha \omega_2 \end{bmatrix}$

$f(\bar{\omega}, \bar{x}) = w_1 x_1 + w_2 x_2$

a) $f(\omega, \bar{x}+\bar{y}) = w_1(x_1+y_1) + \omega_2(x_2+y_2)$

$= w_1 x_1 + w_2 x_2 + w_1 y_1 + w_2 y_2$

$= f(\bar{w}, \bar{x}) + f(\bar{\omega}, \bar{y})$

b) $f(\bar{\omega}, \alpha\bar{x}) = w_1 \alpha x_1 + w_2 \alpha x_2$

$= \alpha(w_1 x_1 + w_2 x_2)$

$= \alpha f(\bar{\omega}, \bar{x})$

↖ Example to elaborate calculations.

Similiarly it can be shown for $\bar{\omega}$

(b) $f(x) = w_1 x_1^2 + w_2 x_2^3$

Linearn in ω

Non-Linear in x

(c) $f(x) = w_1 \ln(x_1) + w_2 e^{x_2}$

Linear in ω

Non linear in x

(d) $f(x) = x_1 \ln(w_1) + x_2 e^{w_2}$

Linear in x

Non Linear in ω

(e) $f(x) = w^T x \quad w, x \in \mathbb{R}^d$

Linear in w and x

(f) $f(x) = w^T x + b$

If we introduce another notation $x_0 = 1$ such that

$\bar{x} = [x_0 \; x_1 \; \ldots \ldots \; x_d]^T$ and call $\omega_0 = b$

then the above equation becomes $f(x) = [w_0 \; w_1 \ldots \ldots w_d] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}$

$f(x) = w^T_{new} x_{new}$ is linear now in ω_{new} and x_{new}

9. L-2 Loss Calculations for 1d case

Note : Wo is used here inplace of b. Dont get confused

$$E = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

$$\frac{\partial E}{\partial w_0} = \sum_{i=1}^{N} 2(y_i - w_0 - w_1 x_i)(-1)$$

Put this to zero we get

$$\sum_{i=1}^{N} (y_i - w_0 - w_1 x_i) = 0$$

$$n\bar{y} - nw_0 - n\bar{x}w_1 = 0$$

$$\boxed{w_0 + \bar{x}w_1 = \bar{y}}$$

$$\boxed{[\bar{x}]w_0 + [\bar{x}^2]w_1 = \bar{x}\bar{y}} \text{ — A}$$

$$\frac{\partial E}{\partial w_1} = \sum_{i=1}^{N} 2(y_i - w_0 - w_1 x_i)(-x_i)$$

Put this to zero we get

$$\sum x_i y_i - w_0(\sum x_i) - w_1(\sum x_i^2) = 0$$

$$\sum x_i y_i - w_0 n\bar{x} - w_1(\sum x_i^2) = 0$$

$$\boxed{[n\bar{x}]w_0 + [\sum x_i^2]w_1 = \sum x_i y_i}$$

$$\boxed{[\bar{x}]w_0 + \left[\frac{\sum x_i^2}{n}\right]w_1 = \frac{\sum x_i y_i}{n}} \text{ — B}$$

from here we get

A − B gives us $\quad w_1 = \dfrac{\sum x_i y_i}{n} - \bar{x}\bar{y} \Big/ \dfrac{\sum x_i^2}{n} - \bar{x}^2$

$$w_0 = \bar{y} - \bar{x}w_1$$

By the way equation of the line was : $y = w_0 + w_1 x$

Putting a bar both sides gives one of the equation

(10) L2 Loss

$$X = \begin{bmatrix} — x_1^T — \\ — x_2^T — \\ — x_3^T — \\ \\ — x_n^T — \end{bmatrix}_{n \times d}$$

By definition design matrix looks like this, where x_i = ith samplevector

Note $x_i \in \mathbb{R}^{d \times 1}$

$x_i^T \in \mathbb{R}^{1 \times d}$

Now by matrix multiplication

$$Xw = \begin{bmatrix} x_1^T w \\ x_2^T w \\ x_3^T w \\ \vdots \\ x_n^T w \end{bmatrix}_{n \times 1} \quad \text{and } Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

$$\|Xw - Y\|^2 = \left\| \begin{bmatrix} x_1^T w - y_1 \\ x_2^T w - y_2 \\ \vdots \\ x_n^T w - y_n \end{bmatrix} \right\|^2 = \sum (x_i^T w - y_i)^2$$

$$= \sum (w^T x_i - y_i)^2$$

$\boxed{\text{NOTE}}$ $x_i^T w = w^T x_i$

Now

$$\mathcal{L}(w) = \|Xw - Y\|^2 = (Xw - Y)^T (Xw - Y) = (w^T X^T - Y^T)(Xw - Y)$$

$$= w^T X^T X w - w^T X^T Y - Y^T X w + Y^T Y$$

we have to find

$$\nabla_w \left[w^T X^T X w - w^T X^T Y - Y^T X w + Y^T Y \right]$$

$$2X^T X w - X^T y - X^T y + 0 = 0$$

$$\left[X^T X w = X^T y \right]$$

Note that a little playing around with matrices will convince you that the above derivatives are correct

this is just an exploratory exercise for you to exercise your Linear Algebraic muscles. Matrix derivatives are a good way to combine multiple equations together

Q 11 Solution

- If A is a full column rank matrix (that is, its columns are independent), $A^T A$ is invertible.
- We will show that the null space of $A^T A$ is $\{0\}$, which implies that the square matrix $A^T A$ is full column (as well as row) rank is invertible. That is, if $A^T A\mathrm{x} = 0$, then $\mathrm{x} = 0$. Note that if $A^T A\mathrm{x} = 0$, then $\mathrm{x}^T A^T A\mathrm{x} = ||A\mathrm{x}|| = 0$ which implies that $A\mathrm{x} = 0$. Since the columns of A are linearly independent, its null space is 0 and therefore, $\mathrm{x} = 0$.

(12) For some $v \neq 0$, $v \in \mathbb{R}^n$ Lets calculate

$$v^T(X^TX + \lambda I)v$$

$$v^TX^TXv + \lambda v^T I v$$

$$(Xv)^T(Xv) + \lambda v^T v$$

$$\|Xv\|^2 + \lambda \|v\|^2$$

unless $v = 0$ $\lambda\|v\|^2$ is positive
$\|Xv\|^2$ is either 0 or positive

therefore

$$v^T(X^TX + \lambda I)v > 0 \quad \forall\, v \neq 0$$

Ie $X^TX + \lambda I$ is positive definite and its inverse always exist

(13) $Y_i \sim \mathcal{N}(w^Tx_i, \sigma^2)$

$$f_{Y_i}(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - w^Tx_i)^2}{2\sigma^2}} \qquad y_i \in (-\infty, \infty)$$

The Likelihood is defined for n iid Y_i's as

$$L(\theta) = \prod_{i=1}^{n} f_{Y_i}(y_i)$$

$\prod \equiv$ product

$\Sigma \equiv$ sum

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - w^Tx_i)^2}{2\sigma^2}}$$

$$L(\theta) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{\sum_1^n (y_i - w^Tx_i)^2}{2\sigma^2}}$$

We are to find the θ that maximises Likelihood
as $\log(\cdot)$ is a monotonically increasing function, that same θ will
also maximise $\log(L(\theta))$

$$\log(L(\theta)) = \underbrace{\log\left(\frac{1}{\sigma^n (2\pi)^{n/2}}\right)}_{\text{constant wrt } \theta} \underbrace{- \frac{\sum_1^n (y_i - w^Tx_i)^2}{2\sigma^2}}_{\text{Negtive of our L-2 loss of Regression}}$$

Hence we have to minimize $\sum_{i=1}^{n}(y_i - w^Tx_i)^2$ which is same as Linear Regression earlier