ECE 421 Homework 2

w is a column vector by default

$f(w) = w^T A$

$\nabla_w f(w) = A$

$f(w) = A^T w$

$\nabla_w f(w) = A$

$f(w) = w^T A w = (A + A^T) w$

Let $x = v^T A$

Question 1,

$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}$, $\frac{\partial f}{\partial w} = x^T = (v^T A)^T = A^T v$

(a) $f(w) = w^T A v + w^T A^T v + v^T A w + v^T A^T w$ → Let $x = v^T A^T$

$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}$

$\frac{\partial f}{\partial w} = x^T = (v^T A^T)^T = A v$

$= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$

$\nabla_w f(w) = A v + A^T v + A^T v + A v$

$= 2Av + 2A^T v$

$= 2(Av + A^T v)$

$= 2(A + A^T) v$

(b) $f(w) = w^T A w$

$\nabla_w f(w) = (A + A^T) w$

(c) $f(w) = \log\left(\sum_{i=1}^{d} e^{w_i}\right)$

$\nabla_w f(w) = \frac{1}{\sum_{i=1}^{d} e^{w_i}} \nabla_w \left(e^{w_1} + e^{w_2} + \cdots + e^{w_d}\right)$

$= \frac{1}{\sum_{i=1}^{d} e^{w_i}} \begin{bmatrix} e^{w_1} \\ e^{w_2} \\ \vdots \\ e^{w_d} \end{bmatrix}$ → Column Vector

(d) $f(w) = \sum_{i=1}^{d} \log(1+e^{w_i}) = \log(1+e^{w_1}) + \log(1+e^{w_2}) + \dots + \log(1+e^{w_d})$

$$\frac{\partial f(w)}{\partial w_1} = \frac{1}{1+e^{w_1}} e^{w_1} = \frac{\partial}{\partial w_1} \log(1+e^{w_1})$$

$$\nabla_w f(w) = \begin{bmatrix} \frac{e^{w_1}}{1+e^{w_1}} \\ \vdots \\ \frac{e^{w_d}}{1+e^{w_d}} \end{bmatrix}$$

(e) $f(w) = \sqrt{1+\|w\|_2^2}$

$= \sqrt{1+w_1^2+w_2^2+\dots+w_d^2}$

$$\frac{\partial f(w)}{\partial w_1} = \frac{1}{2}(1+\|w\|_2^2)^{-\frac{1}{2}} \times 2w_1$$

$$= (1+\underbrace{\|w\|_2^2}_{\text{L2Norm}})^{-\frac{1}{2}} w_1$$

$$\nabla_w f(w) = (1+\|w\|_2^2)^{-\frac{1}{2}} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_d \end{bmatrix}$$

Question 2,

$x_1 = -1, \; y_1 = -2$

$x_2 = 0, \; y_2 = 0$

$x_3 = 1, \; y_3 = 1$

$$E_{in}(w) = \frac{1}{3}\sum_{i=1}^{3}(y_i - \hat{y}_i)^2$$

$$\hat{y}_i = w_0 + w_1 x_i$$

a) $$w^* = \underset{w \in \mathbb{R}^2}{\operatorname{argmin}} \; \frac{1}{3}\|\mathcal{X}w - \underline{y}\|^2$$

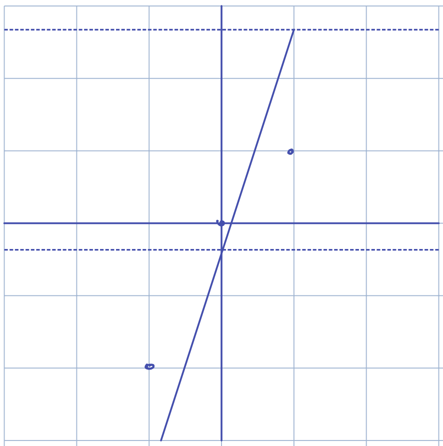
$$\underline{\mathcal{X}} = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

(b) $$w^* = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\underline{y}$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ \frac{3}{2} \end{bmatrix}$$



$$\hat{y}_i = -\frac{1}{3} + \frac{3}{2}x_i$$

(c) $\hat{y} = w_0 + w_1 x + w_2 x^2$

$$\mathcal{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$E_{in}(w)$

$$w^* = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \underline{y}$$

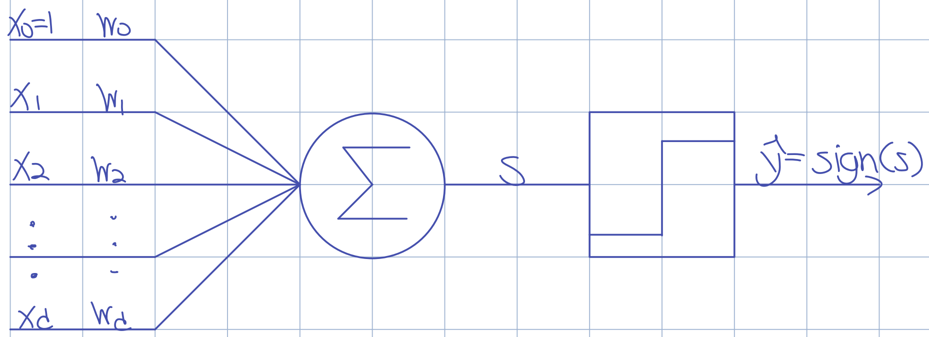
$$= \begin{bmatrix} 0 \\ \frac{3}{2} \\ -\frac{1}{2} \end{bmatrix}$$

$$E_{in}(w^*) = \frac{1}{3} \left\| \mathcal{X} w - y \right\|^2$$

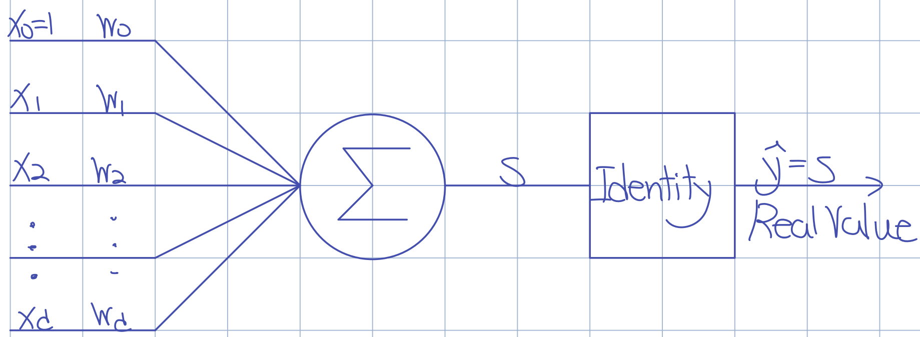
$$= \frac{1}{3} \left\| \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ \frac{3}{2} \\ -\frac{1}{2} \end{bmatrix} - \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\|^2$$

$$= 0$$

Binary Linear Classification (Hard Decision -1 or 1)

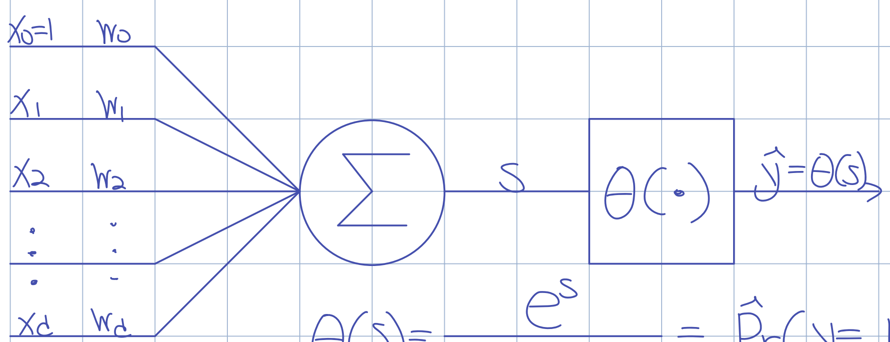


Linear Regression



Logistic Regression

Soft Decision Between -1 and 1



$$\theta(s) = \frac{e^s}{1+e^s} = \hat{\Pr}(y=1 \mid x)$$

$$\hat{P}_w(1 \mid x) = \theta(w^T x) = \frac{e^{w^T x}}{1+e^{w^T x}}$$

$$\hat{P}_w(-1 \mid x) = 1 - \hat{P}_w(1 \mid x) = 1 - \frac{e^{w^T x}}{1+e^{w^T x}}$$

$$= \frac{1}{1+e^{w^T x}} = \frac{e^{-w^T x}}{1+e^{-w^T x}}$$

$$\therefore \hat{P}_w(y \mid x) = \frac{e^{y w^T x}}{1+e^{y w^T x}} = \frac{1}{1+e^{-y w^T x}}$$

Loss Function (Log Loss)

$$Loss = -\log \hat{P}_w(y \mid x) = -\log\left(\frac{1}{1+e^{-y w^T x}}\right) = \log(1+e^{-y w^T x})$$

y_n	$w^T x_n$	$y_n w^T x_n$	$e_n(w)$
$+1$	$\gg 0$	$\gg 0$	Small
$+1$	$\ll 0$	$\ll 0$	Large
-1	$\gg 0$	$\ll 0$	Large
-1	$\ll 0$	$\gg 0$	Small

$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} \log(1+e^{-y w^T x})$$

$$w^* = \underset{w}{\operatorname{argmin}}\, E_{in}(w)$$

Regularization

$$\min\{E_{in}(w) + \lambda \|w\|^2\}$$ L1 Norm

Maximum Likelihood

$$\Pr(y_1, y_2, \dots, y_N \mid x_1, x_2, \dots, x_N) = \prod_{n=1}^{N} \hat{\Pr}(y_n \mid x_n)$$

$$w^* = \underset{w}{\operatorname{argmax}} \prod_{n=1}^{N} \hat{\Pr}(y_n \mid x_n)$$

$$= \underset{w}{\operatorname{argmin}}\, E_{in}(\underline{w})$$

Part A,

$y = +1 \longrightarrow P(y_n \mid x_n) = h(x_n)$

$y = -1 \longrightarrow P(y_n \mid x_n) = 1 - h(x_n)$

$$\max \prod_{n=1}^{N} \hat{\Pr}(y_n \mid x_n) = \max \ln \prod_{n=1}^{N} \hat{\Pr}(y_n \mid x_n)$$

$$= \max \sum_{n=1}^{N} \ln \hat{\Pr}(y_n \mid x_n)$$

$$= \min \sum_{n=1}^{N} - \ln \hat{\Pr}(y_n \mid x_n)$$

$$E_{in}(w) = - \sum_{n=1}^{N} \ln \hat{\Pr}(y_n \mid x_n)$$

$$= - \sum_{n=1}^{N} \underbrace{[\![y_n = +1]\!]}_{I(y=+1)} \ln h(x_n) + \underbrace{[\![y_n = -1]\!]}_{I(y=-1)} \ln(1 - h(x_n))$$

$$= \sum_{n=1}^{N} [[y_n = +1]] \ln \frac{1}{h(x_n)} + [[y_n = -1]] \ln \frac{1}{1-h(x_n)}$$

Part B,

$$\theta(w^T x) = \frac{e^{w^T x}}{1+e^{w^T x}} = \frac{1}{1+e^{-w^T x}}$$

$$E_{in}(w) = \sum_{n=1}^{N} [[y_n = +1]] \ln \frac{1}{h(x_n)} + [[y_n = -1]] \ln \frac{1}{1-h(x_n)}$$

$$\frac{1}{h(x_n)} = 1+e^{-w^T x}$$

$$\frac{1}{1-h(x_n)} = 1+e^{w^T x}$$

$$\boxed{\begin{array}{l} y_n = 1, \ \ln(1+e^{-w^T x}) \\ y_n = -1, \ \ln(1+e^{w^T x}) \end{array}}$$

$$E_{in}(w) = \sum_{n=1}^{N} [[y_n = +1]] \ln(1+e^{-w^T x}) + [[y_n = -1]] \ln(1+e^{w^T x})$$

$$= \sum_{n=1}^{N} \ln(1+e^{-y_n w^T x})$$

Exercise 3.7,

$$\nabla E_{in}(w) = \nabla \frac{1}{N} \sum_{n=1}^{N} \ln(1+e^{-y_n w^T x_n})$$

$$= \frac{1}{N} \nabla \sum_{n=1}^{N} \ln(1+e^{-y_n w^T x_n})$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{-y_n x_n e^{-y_n w^T x_n}}{1+e^{-y_n w^T x_n}} = \frac{-1}{N} \sum_{n=1}^{N} \frac{y_n x_n}{1+e^{y_n w^T x_n}}$$

$$= \frac{-1}{N} \sum_{n=1}^{N} y_n x_n \frac{e^{-y_n w^T x_n}}{1 + e^{-y_n w^T x_n}}$$

$$= \frac{-1}{N} \sum_{n=1}^{N} y_n x_n \, \theta(-y_n w^T x_n)$$

Miss Classification, $y_n w^T x_n < 0$, $\theta(+) > 0.5$

Correct Classification, $y_n w^T x_n > 0$, $\theta(-) < 0.5$