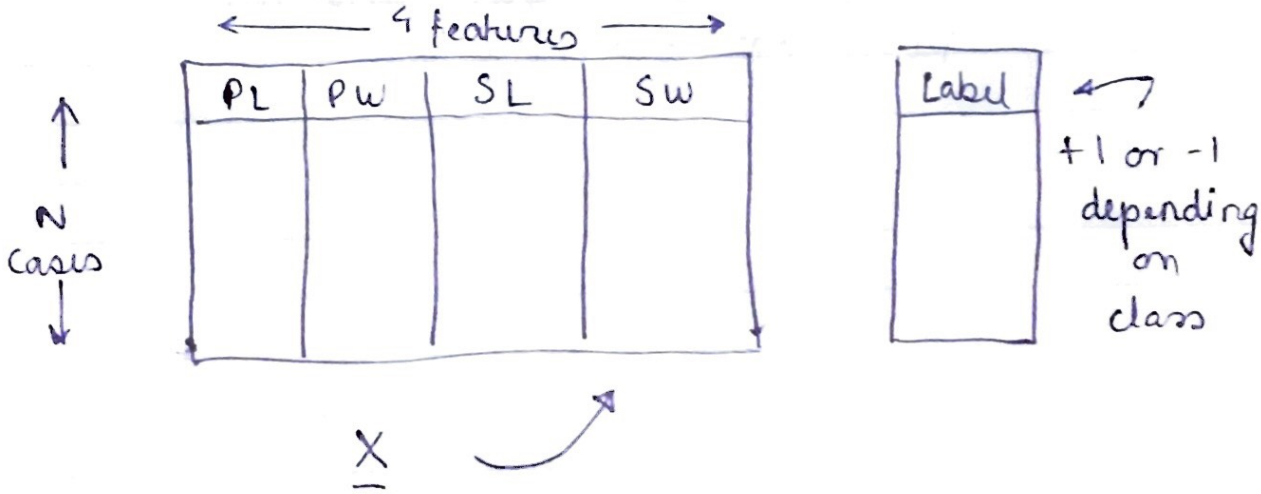
EE-660 ASSIGNMENT 1

Tanwoghura

Q1 a) $\mathcal{D} = \{ \underline{x}_i, y_i \}_{i=1}^{N}$

2 types : Virginica $(y = +1)$
Versicolor $(y = -1)$

4 features : petal length, petal width, sepal length, sepal width



$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad \underline{\underline{X}} = \begin{bmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_N^T & — \end{bmatrix}$$

output vector — design matrix

b) The dimension of $\underline{\underline{X}}$ is $\boxed{N \times 4}$

c) Two features : petal length and petal width.

$x_1 \longrightarrow$ petal length
$x_2 \longrightarrow$ petal width

Then, linear classifier :

$$\hat{y}(\underline{x}) = w_0 + w_1 x_1 + w_2 x_2$$

Referring to Fig 1.4 for the dataset, a linear classifier will be able to classify most of the dataset correctly given the two features petal length and petal width.

Q2

$$\hat{f}(x) = \sum_{k=0}^{d} w_k x^k$$

d - given

x - time

$1 \le x_i \le 365$

$x_i \in \mathbb{Z}$ ——— ①

(a) $\Phi(x) = [1 \; x \; x^2 \; x^3 \; \dots \; x^d]$

dimension of $\Phi(x)$ = $(d+1) \cdot 1$

(b) The weight vector $\underline{w}$ which includes the variables w_k's in the above equation ① are to be found using a learning algorithm.

Q3. (a) The parameter vector to be estimated $\underline{\theta}$, in the notation of our stock price problem, i.e., equation ① is the weight vector $\underline{w}$

(b) Murphy assumes that the probability distribution $p(y|x,\underline{\theta})$ is Normal / Gaussian distribution given by,

$$N(y \mid \mu(x), \sigma^2(x))$$

where assumption is, $\mu = w^T x$

noise is fixed

$\sigma^2(x) = \sigma^2$

$\theta = (w, \sigma^2)$

The Training samples are commonly assumed to be independent and identically distributed (iid).

Q4. MAP estimation:

$$\underset{w}{\operatorname{argmax}} \sum_{i=1}^{N} \log N(y_i \mid w_0 + w^T x_i, \sigma^2) + \sum_{j=1}^{D} \log N(w_j \mid 0, \tau$$

$$= \underset{w}{\operatorname{argmax}} \sum_{i=1}^{N} \log \left[\left(\frac{1}{2\pi\sigma^2}\right)^{1/2} e^{\left(-\frac{1}{2\sigma^2} (y_i - (w_0 + w^T x_i))^2\right)} \right]$$

$$+ \sum_{j=1}^{D} \log \left[\left(\frac{1}{2\pi\tau^2}\right)^{1/2} e^{\left(\frac{-1}{2\tau^2} (w_j)^2\right)} \right]$$

$$= \underset{w}{\operatorname{argmax}} \sum_{i=1}^{N} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_i - (w_0 + w^T x_i))^2 \right]$$

$$+ \sum_{j=1}^{D} \left[-\frac{1}{2} \log(2\pi\tau^2) - \frac{1}{2\tau^2} (w_j)^2 \right]$$

Here, $-\frac{1}{2}\log(2\pi\sigma^2)$ and $-\frac{1}{2}\log(2\pi\tau^2)$ are constant terms so, we can remove them and then the equation becomes,

$$= \underset{w}{\operatorname{argmax}} \sum_{i=1}^{N} -\frac{1}{2\sigma^2} (y_i - (w_0 + w^T x_i))^2$$

$$+ \sum_{j=1}^{D} \frac{-1}{2\tau^2} (w_j)^2$$

$$= \underset{w}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2\sigma^2} (y_i - (w_0 + w^T x_i))^2$$

$$+ \sum_{j=1}^{D} \frac{1}{2\tau^2} (w_j)^2$$

Here we changed from argmax to argmin, hence the sign change.

$$= \underset{w}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2\sigma^2} \left(y_i - (w_0 + w^T x_i) \right)^2$$

$$+ \frac{1}{2\tau^2} \|w\|_2^2$$

$$\because \quad \Sigma_j w_j^2 = w^T w = \|w\|_2$$

Multiplying with constant $2\sigma^2$,

$$= \frac{1}{2\sigma^2} \underset{w}{\operatorname{argmin}} \sum_{i=1}^{N} \left(y_i - (w_0 + w^T x_i) \right)^2 + \lambda \|w\|_2^2$$

where,

$$\lambda \triangleq \frac{\sigma^2}{\tau^2}$$

As $\frac{1}{2\sigma^2}$ is a constant, we can ignore it. So,

$$\boxed{J(w) = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{N} \left(y_i - (w_0 + w^T x_i) \right)^2 + \lambda \|w\|_2^2}$$