Theory Questions

Assignment 4.

1.

a) $E_D(w) = \frac{1}{2} \sum_{n=1}^{N} g_n \left(t_n - w^T \phi(x_n)\right)^2$

To find w which minimizes,

$$\nabla E_D(w) = \frac{1}{2} \times 2 \sum_{n=1}^{N} g_n \left(t_n - w^T \phi(x_n)\right) * \phi(x_n)^T$$

let

$$\sqrt{g_n}\, \phi(x_n) = \phi'(x_n) \quad \& \quad \sqrt{g_n}\, t_n = t'_n$$

$$\Rightarrow \sum_{n=1}^{N} g_n t_n \phi(x_n)^T - \sum_{n=1}^{N} g_n w^T \phi(x_n)\, \phi(x_n)^T = 0$$

$$\sum_{n=1}^{N} \phi'(x_n)^T t'_n - \sum_{n=1}^{N} w^T * \phi'(x_n) * \phi'(x_n)^T = 0$$

$$\sum_{n=1}^{N} \phi'(x_n)^T t'_n = w^T \left(\sum_{n=1}^{N} \phi'(x_n) * \phi'(x_n)^T \right)$$

From this equations

$$\boxed{w^* = (\phi^T \phi)^{-1} * \phi^T . t}$$

$$t = \left[\sqrt{g_1}\, t_1, \sqrt{g_2}\, t_2, \ldots \sqrt{g_N}\, t_N\right]^T$$

We also define ϕ as a $N \times M$ matrix, with element

$\phi(i,j) = \sqrt{r_i}\, \phi_j(x_i)$.

Interpretation :-

i) Data dependent Noise Variance:-

Considering a guassian noise model, let us assume target variable t is given by deterministic function $y(x,w)$ with additive Guassian noise so that

$$t = y(x,w) + \varepsilon$$

Where ε is zero mean Guassian RV with precision (inverse variance) β.

$$p(t|x,w,\beta) = \mathcal{N}(t|y(x,w), \beta^{-1})$$

$$\mathbb{E}(t|x) = \int t\,p(t|x)\cdot dt$$

$$= y(x,w)$$

$$p(t|x,w,\beta)$$

$$= \prod_{i=1}^{n} \mathcal{N}(t_n|w^T\phi(x_n), \beta^{-1})$$

Consider log

$$\ln p(t|N,\beta) = \sum_{n=1}^{N} \ln(\mathcal{N}(t_n|w^T\phi(x_n)\,\beta^{-1})$$

$$= \frac{N}{2}\ln\beta - \frac{N}{2}\ln 2\pi - \beta E_D(w)$$

SSE is

$$E_D(w) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - w^T\phi(x_n)\}^2$$

Consider gradient

$$\nabla \ln p(t|w,\beta) = \sum_{n=1}^{N}\{t_n - w^T\phi(x_n)\}\phi(x_n)^T$$

Setting gradient to 0

$$0 = \sum_{n=1}^{N} t_n\phi(x_n)^T - w^T\sum_{n=1}^{N}\phi(x_n)\phi(x_n)^T$$

$$\Rightarrow \boxed{w = (\phi^T\phi)^{-1}\phi^T t} \rightarrow (2).$$

ii) From given question g_n can also be viewed as effective number of observation (x_n, t_n) i.e (x_n, t_n) can be treated as repeatedly occuring g_n times [replicate data points].

② Bayes estimate

A: $\sum_{i=1}^{5} P(h_i/D) * P(F/h_i) = 0.4$

B: $\sum_{i=1}^{5} P(h_i/D) * P(L/h_i) = 0.2\times1+0.1\times1+0.2\times1$
$= 0.5$

C: $\sum_{i=1}^{5} P(h_i/D) * P(R/h_i) = 0.1\times1$
$= 0.1$

$B > A > C$ So Robot moves left.

MAP.

Max hypothesis is given as

$\hat{h} = \arg\max P(h/D)$

With given data $h_1 = 0.4$ is highest

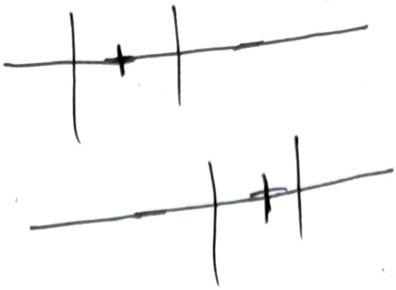
With h_1 $P(F/h_1) = 1$ $P(L/h_1) = 0$ $P(R/h_1) = 0$

So Robot moves Forward with MAP rule

3. One-d data So data will be on straight line

Consider 2 points

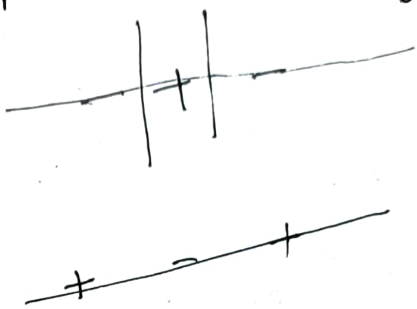
$p < x < q$



all different labels of 2 points is satisfied

Consider 3 points

✓



✗ Can't be done

So $vc(H) = 2$.

4. We arrange $E_D(w)$

$$E_D(w) = \frac{1}{2}\sum_{n=1}^{N}\left\{\overbrace{\left[w_0 + \sum_{i=1}^{D} w_i(x_i+\epsilon_i) - t_n\right.}^{y_i}\right\}^2$$

$$= \frac{1}{2}\sum_{n=1}^{N}\left\{\left(w_0 + \sum_{i=1}^{D} w_i x_i\right) - t_n + \sum_{i=1}^{D} w_i\epsilon_i\right\}^2$$

$$= \frac{1}{2}\sum_{n=1}^{N}\left\{y(x_n, w) - t_n + \sum_{i=1}^{D} w_i\epsilon_i\right\}^2$$

$$= \frac{1}{2}\sum_{n=1}^{N}\left\{(y(x_n, w) - t_n)^2 + \left(\sum_{i=1}^{D} w_i\epsilon_i\right)^2 + 2\left(\sum_{i=1}^{D} w_i\epsilon_i\right)(y(x_n, w) - t_n)\right\}$$

Where we have used $y(x_n, w)$ to denote the output of the linear model when input variable is x_n, without noise added. For the second term in the equation above, we can obtain:

$$\mathbb{E}_\epsilon\left[\left(\sum_{i=1}^{D} w_i \epsilon_i\right)^2\right] = \mathbb{E}_\epsilon\left[\sum_{i=1}^{D}\sum_{j=1}^{D} w_i w_j \epsilon_i \epsilon_j\right]$$

$$= \sum_{i=1}^{D}\sum_{j=1}^{D} w_i w_j \mathbb{E}_\epsilon[\epsilon_i \epsilon_j]$$

$$= \sigma^2 \sum_{i=1}^{D}\sum_{j=1}^{D} w_i w_j \delta_{ij}$$

Which gives

$$\mathbb{E}_\epsilon\left[\left(\sum_{i=1}^{D} w_i \epsilon_i\right)^2\right] = \sigma^2 \sum_{i=1}^{D} w_i^2$$

For the third term, we can obtain:

$$\mathbb{E}_\epsilon\left[2\left(\sum_{i=1}^{D} w_i \epsilon_i\right)(y(x_n, w) - t_n)\right]$$

$$= 2\left[y(x_n, w) - t_n\right] \mathbb{E}_\epsilon\left[\sum_{i=1}^{D} w_i \epsilon_i\right]$$

$$= 2(y(x_n, w) - t_n) \sum_{i=1}^{D} [w_i \epsilon_i]$$

$$= 0$$

Therefore, if we calculate the expectation of $E_D(w)$ with respect to ϵ, we can obtain

$$\mathbb{E}_\epsilon[E_D(w)] = \frac{1}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2 + \frac{\sigma^2}{2}\sum_{i=1}^{D} w_i^2$$