Q1) Distance from mean.

In this given a test example we say the label is $y = \text{sign}(f(x))$. → x feature vector of test

where → y predicted label.

$$f(x) = \|\mu_- - x\|^2 - \|\mu_+ - x\|^2$$

$$= \underbrace{\mu_-^2 - \mu_+^2}_{b} + 2\mu_+^T x - 2\mu_-^T x.$$

$$= b + 2\left[\mu_+ - \mu_-\right]^T x$$

$$= b + 2\left[\sum_{n \in N_+} \frac{x_n}{|N_+|} - \sum_{n \in N_-} \frac{x_n}{|N_-|}\right]^T x.$$

$$= \sum_{n=1}^{N} \alpha_n \langle x_n, x \rangle + b$$

where $\alpha_n = \begin{cases} \dfrac{2y_n}{|N_+|} & \text{if } y_n > 0 \\ \dfrac{2y_n}{|N_-|} & \text{if } y_n < 0 \end{cases}$

$b = \mu_-^2 - \mu_+^2$

hence we get the form

$$f(x) = \sum_{n=1}^{N} \alpha_n \langle x_n, x \rangle + b$$

Q2) Gausian Distribution

$$N(x) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left[-1/2\,(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$$

taking the ration

$$\frac{N_+(x)}{N_-(x)}$$

Now we want a function f such that label of a testing feature vector x is given by

$$y = \text{sign}(f(x))$$

Define

$$f(x) = \ln \frac{N_+(x)}{N_-(x)}.$$

$$= \ln \left[\frac{\frac{1}{\sqrt{(2\pi)^{D_+}}\sqrt{|\Sigma|}} \exp\left[-1/2\,(x-\mu_+)^T \Sigma^{-1} (x-\mu_+)\right]}{\frac{1}{\sqrt{(2\pi)^{D_-}}\cdot\sqrt{|\Sigma|}} \exp\left[-1/2\,(x-\mu_-)^T \Sigma^{-1} (x-\mu_-)\right]} \right]$$

$$= \cancelto{0}{\frac{1}{2} \ln \frac{(2\pi)^{D_+}}{(2\pi)^{D_-}}} - 1/2\,(x-\mu_+)^T \Sigma^{-1} (x-\mu_+) + \frac{1}{2}(x-\mu_-)^T \Sigma^{-1} (x-\mu_-)$$

$$= \cancelto{0}{\frac{D_+ - D_-}{2} \ln 2\pi} + 1/2 \left[(x^T\Sigma^{-1} - \mu_-^T \Sigma^{-1})(x-\mu_-) - (x^T\Sigma^{-1} - \mu_+^T\Sigma^{-1})(x-\mu_+) \right]$$

$(\because$ using $x^T\Sigma^{-1}\mu_- = \mu_-^T\Sigma^{-1}x)$

$$= \underbrace{\cancelto{0}{\frac{D_+ - D_-}{2} \ln 2\pi} + 1/2\left[\mu_-^T\Sigma^{-1}\mu_- - \mu_+^T\Sigma^{-1}\mu_+\right]}_{b} + \underbrace{\left[\mu_+^T\Sigma^{-1} - \mu_-^T\Sigma^{-1}\right]}_{w^T} x$$

$$= b + w^T x$$

$D_+ = D_-$
since D is gaussian distribution is the dimension.

$$w = (\mu_+^T\Sigma^{-1} - \mu_-^T\Sigma^{-1})^T$$

$$b = \cancelto{0}{\frac{D_+ - D_-}{2} \ln 2\pi} + 1/2\left[\mu_-^T\Sigma^{-1}\mu_- - \mu_+^T\Sigma^{-1}\mu_+\right]$$

Q3) linear regression

We change then Lemp to incorporate the weight cn

$$L_{emp} = 1/2 \sum_{n=1}^{N} (y_n - w^T x_n)^2 c_n$$

$$\Rightarrow \frac{dL_{emp}}{dw} = \sum_{n=1}^{N} (y_n - w^T x_n) \cdot c_n x_n$$

Equating $\frac{dL_{emp}}{dw}$ to zero

$$\sum_{n=1}^{N} y_n c_n x_n = \left(\sum_{n=1}^{N} w^T x_n\right) \cdot c_n x_n$$

$$\Rightarrow \sum_{n=1}^{N} y_n c_n x_n = \left(\sum_{n=1}^{N} x_n^T w\right) c_n x_n.$$

$$\Rightarrow \sum_{n=1}^{N} y_n c_n x_n = \left[\sum_{n=1}^{N} c_n x_n x_n^T\right] w.$$

$$w = \left[\sum_{n=1}^{N} c_n x_n x_n^T\right]^{-} \left[\sum_{n=1}^{N} y_n c_n x_n\right]$$

$$= [X^T X]^{-} [X^T Y]$$

$$X = [\sqrt{c_1} x_1, \sqrt{c_2} x_2 \dots \sqrt{c_n} x_n]^T$$

$$Y = [\sqrt{c_1} y_1, \sqrt{c_2} y_2 \dots \sqrt{c_n} y_n]^T$$

Q4) Noise as reguliser.

$$L(w) = \sum_{n=1}^{N} (y_n - w^T x_n)^2$$

Now replacing x_n with $x_n + \varepsilon_n$

$$\tilde{L}(w) = \sum_{n=1}^{N} (y_n - w^T(x_n + \varepsilon_n))^2$$

Also

$$E(\varepsilon_n) = 0 \qquad E(\varepsilon_n \varepsilon_m) = \delta_{mn} \sigma^2 I$$

$$E(\tilde{L}(w)) = E\left[\sum_{n=1}^{N} (y_n - w^T(x_n + \varepsilon_n))^2\right]$$

$$= E\left[\sum_{n=1}^{N} \left[y_n^2 + (w^T(x_n + \varepsilon_n))^2 - 2y_n w^T(x_n + \varepsilon_n)\right]\right]$$

$$= E\left[\sum_{n=1}^{N} \left[y_n^2 + (w^T x_n)^2 + (w^T \varepsilon_n)^2 - 2(w^T x_n)(w^T \varepsilon_n) - 2y_n w^T(x_n + \varepsilon_n)\right]\right]$$

$$= \sum_{n=1}^{N} y_n^2 + (w^T x_n)^2 + 2w^T x_n \cancel{(w^T E(\varepsilon_n))} + w^T E(\varepsilon_n \varepsilon_n^T) w - 2y_n w^T x_n - 2y_n \cancel{w^T E(\varepsilon_n)}$$

$$= \sum_{n=1}^{N} y_n + (w^T x_n)^2 + w^T E(\varepsilon_n \varepsilon_n^T) w - 2y_n w^T x_n$$

$$= \sum_{n=1}^{N} \left[(y_n - w^T x_n)^2 + \sigma^2 w^T w\right]$$

$$= \sum_{n=1}^{N} (y_n - w^T x_n)^2 + N\sigma^2 w^T w$$

$\Rightarrow E(\tilde{L}(w))$ is the same as the regularized objective function for linear regression.

Scanned by CamScanner

Q5) Decision tree for Regression

Say the proble has n attributes $x_1, x_2 \dots x_n$
Now for the decision tree we need to split the dataset based on some attribute x_i at some value k. So to find x_i and k we use the following method.

Define

$$D = \sum_{x_{ij} > k} (\mu_1 - y_j)^2 + \sum_{x_{ij} < k} (\mu_2 - y_j)^2$$

Note my nomenclature - x_{ij} is the value of x_i feature for j^{th} example.

μ_1 = average over all y such that the corresponding feature x_i for that y is greater than k.

μ_2 = average over all y such that the corresponding feature x_i for that y is less than k.

→ We are findinding a standard deviation like measure

→ We find D for different values of i and k and take that pair of (i, k) for which D is minimum

→ keep doing process recursively.

→ finding condition - keep doing this until you get perfect branches or you can keep doing it for fixed number of iterations.

Q 6).

Initially the accuracy increases rapidly by increasing so more training data but after one more the accuracy does not change so much.

This is mostly because initially on add so more training data the mean of each class change a lot but after some time the mean moves by very small distance and so we don't have much change in the accuracy.