net id = axr8193

Machine learning assignment

1. a) optimization function: $y_i(\omega^T x_i + b) \geq 1$
$i = 1, \ldots, n$

$$\min_{\gamma, \omega, b} \frac{1}{2} ||\omega||^2$$

WKT,

Lagrange dual $\theta_D(\alpha, \beta) = \min_{\omega} L(\omega, \alpha, \beta)$

Re-writing the constraints

$$g(\omega) = y_i(\omega^T x_i + b) - 1 \leq 0$$

When we substitute in our Lagrangian formulla we get:

$$f(\omega) = \frac{1}{2} ||\omega||^2 \quad \left(\max \frac{1}{||\omega||} \approx \min ||\omega|| \approx \min \frac{1}{2} ||\omega||^2 \right)$$

$$g_i(\omega) = y_i(\omega^T x_i + b) - 1$$

$$L(\omega, \alpha, \beta) = \frac{1}{2} ||\omega||^2 + \sum_{i=1}^{m} \alpha_i \left[y_i (\omega^T x_i + b) - 1 \right]$$

$$\boxed{\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^{m} \alpha_i y_i x_i}$$

Calculating for extremum we get

$$w - \sum_{i=1}^{m} \alpha_i y_i x_i = 0$$

$$\boxed{w = \sum_{i=1}^{m} \alpha_i y_i x_i}$$

diff. w.r.t b

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{m} \alpha_i y_i$$

Solving for extremum we get:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{m} \alpha_i y_i = 0$$

Placing them into Lagrange formula:

$$L(w, \alpha, b) = \frac{1}{2} \left(\sum_{i=1}^{m} \alpha_i y_i \bar{x}_i\right) \cdot \left(\sum_{j=1}^{m} \alpha_j y_j \bar{x}_j\right) - \sum_{i=1}^{m} \alpha_i y_i \times \left(\sum_{j=1}^{m} \alpha_j y_j \bar{x}_j\right)$$

$$= \qquad - \underbrace{\cancel{\sum \alpha_i y_i b}}_{0} + \sum \alpha_i$$

$$\boxed{L(w, \alpha, b) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j (\bar{x}_i)^T \cdot \bar{x}_j}$$

$$\Rightarrow \max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle$$

$$\alpha_i \geq 0 \quad i = 1 \ldots n$$

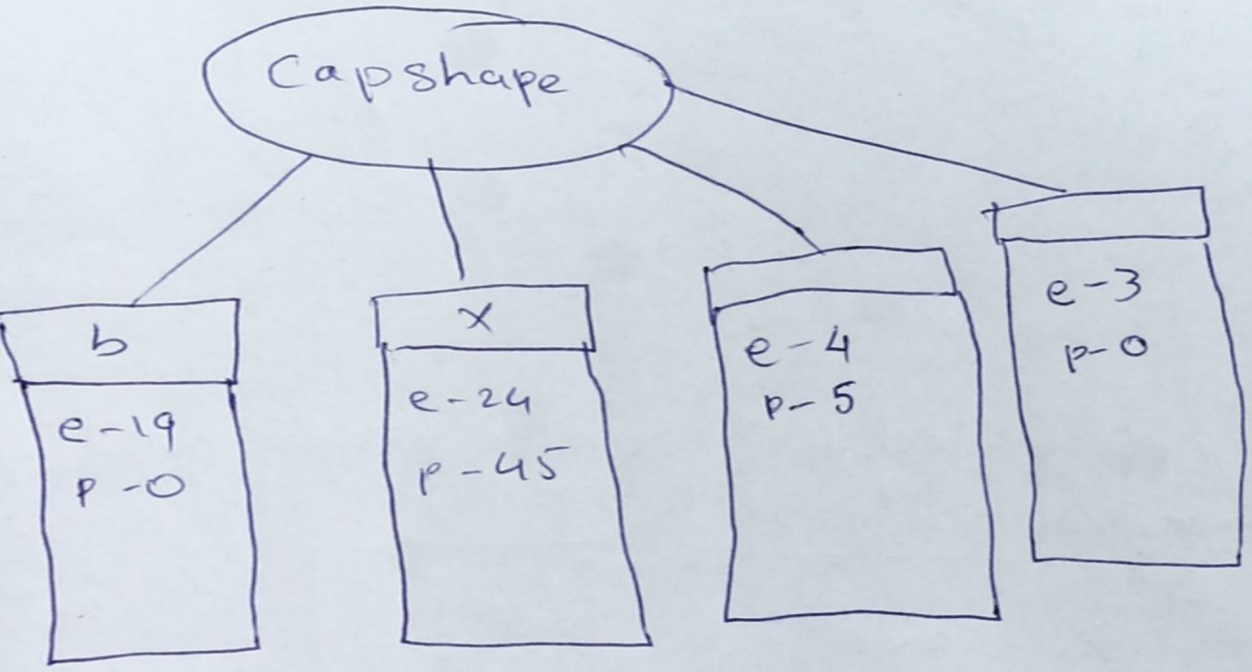
$$\boxed{\sum_{i=1}^{m} \alpha_i y_i = 0}$$

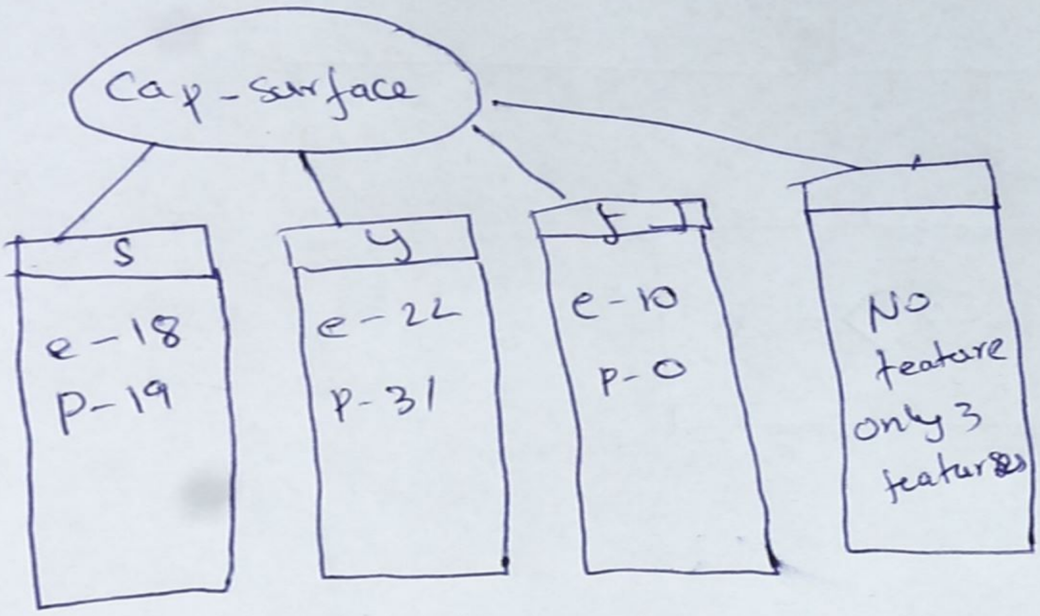
Once a^* has been calculated w^* can be calculated from the dual function with b:

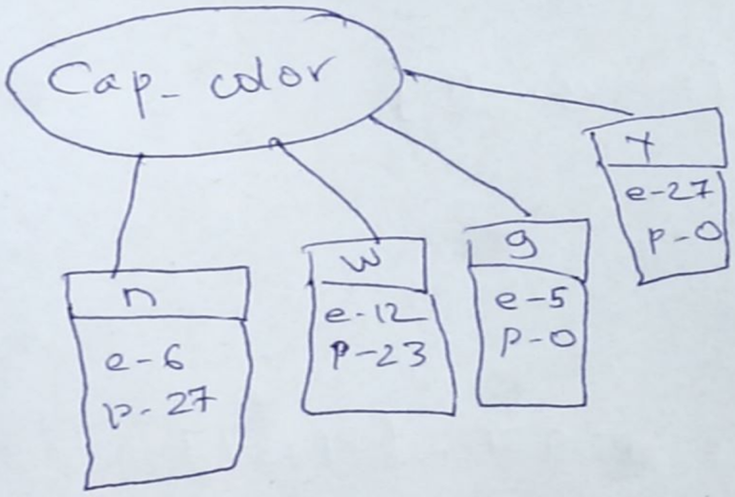
$$\boxed{b = \frac{-\max_{i:y_i=-1} w^{*T} x_i + \min_{i:y_i=1} w^{*T} x_i}{2}}$$

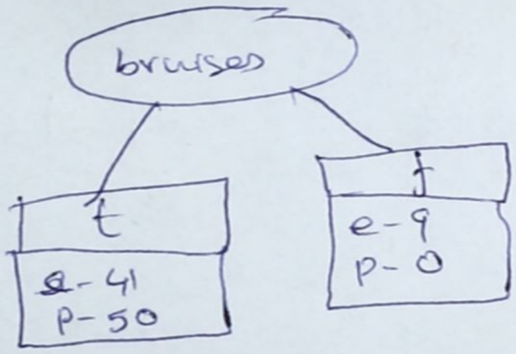
$$\Rightarrow \boxed{w^T x + b = \sum_{i=1}^{m} \alpha_i y_i \langle x^{(i)}, x \rangle + b}$$

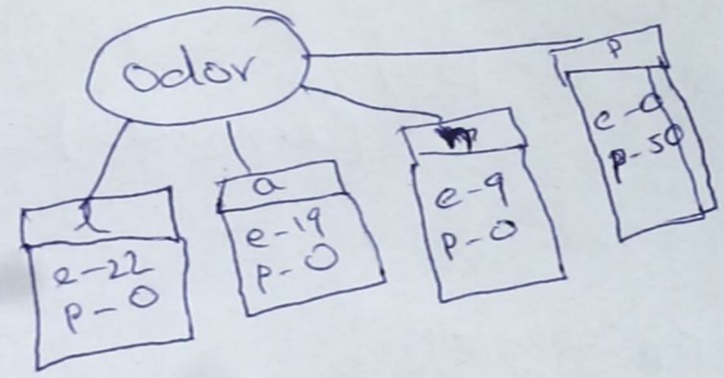
2. a Writing down number of e's and p's for each feature we get











$$E(S) = -p(e)\log_2 p(e) - p(p)\log_2 p(p)$$
$$= -0.5 \times \log_2 0.5 - 0.5 \log_2 0.5$$
$$= 1$$

$$E(capshape=b) = -p(e)\log_2 p(e) - P(p)\log_2 P(p) \rightarrow ①$$
$$= -1 \log_2 1 - 0 \log_2 0$$
$$= 0$$

$$E(capshape=x) = -(24/69) \times -1.523 - (45/69) \times -0.6168 \quad [\text{Applying } ①]$$
$$= 0.9319$$

$$E(capshape=f) = -(4/9) \times -1.173 - (5/9) \times -0.85 \quad [\text{Applying } ①]$$
$$= 0.9918$$

bruise

$$E(bruise = t) = -\left(\frac{41}{91}\right)(-1.1504) - \left(\frac{50}{91}\right)(-0.86407)$$

[Applying ①]

$$= 0.99307$$

$$E(bruise = f) = -1 \cdot 0 - 0$$
$$= 0$$

$E(odor = \text{all } \cancel{\text{features}} \text{ properties}) = 0$ [odor is a pure feature]

Information of capshape

$$I(capshape) = 0 + \frac{69}{100} \times 0.9319 + \frac{9}{100} \times 0.9918 + 0$$

$$= \sum_{i=1}^{n=3} p(\text{first property}) \times E(\text{property}) \rightarrow ②$$

Information of capsurface :

$$I(\text{cap-surface}) = \frac{37}{100} \times 0.9998 + \frac{53}{100} \times 0.98016 \text{ [Applying (2)]}$$

$$= 0.8894$$

$$I(\text{cap-color}) = \frac{33}{100} \times 0.6835 + \frac{35}{100} \times 0.9276 + 0 + 0$$

[Applying (2)]

$$= 0.5502$$

Similarly Applying (2) for the next two features we get:

$$I(\text{bruise}) = \frac{91}{100} \times 0.99307 + 0$$

$$= 0.9036$$

$$I(\text{odour}) = 0$$

After this we find the Information gained from each feature:

$$\text{Gained}(\text{capshape}) = E(S) - I(\text{capshape}) \rightarrow (3)$$

$$= 1 - 0.732273$$

$$= 0.267727$$

Applying (3) we get:

$$\text{Gained}(\text{cap-Surface}) = 1 - 0.8894$$

$$= 0.1106$$

$$I(\text{cap-surface}) = \frac{37}{100} \times 0.9998 + \frac{53}{100} \times 0.98016 \text{ [Applying ②]}$$

$$= 0.8894$$

$$I(\text{cap-color}) = \frac{33}{100} \times 0.6835 + \frac{35}{100} \times 0.9276 + 0 + 0$$

[Applying ②]

$$= 0.5502$$

Similarly Applying ② for the next two features we get:

$$I(\text{bruise}) = \frac{91}{100} \times 0.99307 + 0$$

$$= 0.9036$$

$$I(\text{odour}) = 0$$

After this we find the Information gained from each feature:

$$\text{Gained}(\text{capshape}) = E(S) - I(\text{capshape}) \rightarrow ③$$

$$= 1 - 0.732273$$

$$= 0.267727$$

Applying ③ we get:

$$\text{Gained}(\text{cap-Surface}) = 1 - 0.8894$$

$$= 0.1106$$

Gained (cap. color) = 1 − 0.5502
= 0.4498

Gained (bruise) = 1 − 0.9036
= 0.0964

Gained (odor) = 1 − 0
= 1

Since odor feature has the highest Information gain odor becomes the root.

The below tree is as follows.