Ans 1 (a)

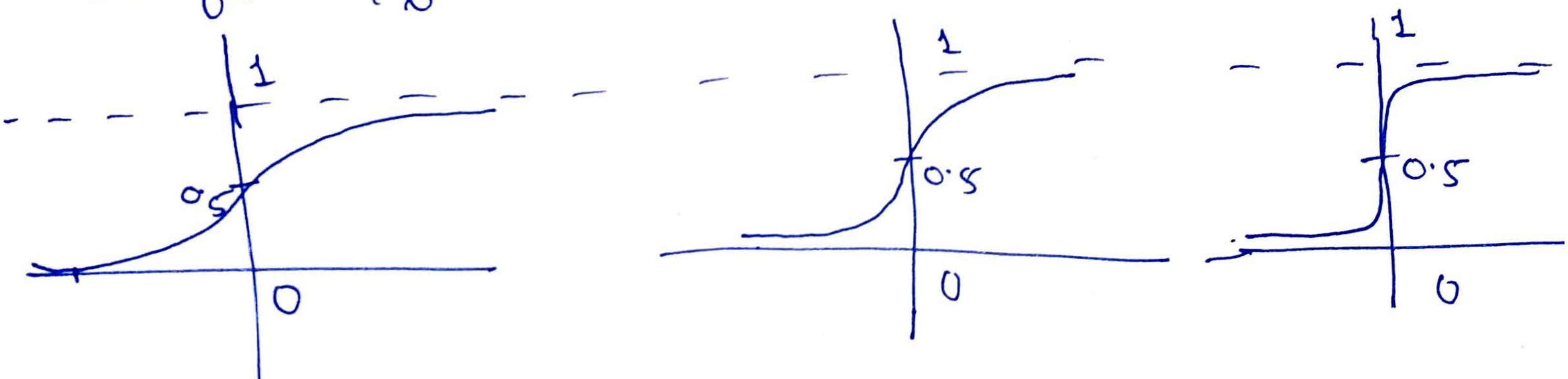
$x \in R$ $\qquad$ $w \in \{1, 5, 100\}$

$$\text{Sigmoid } f^n = \frac{1}{(1+e^{-xw})}$$

Based of w the shape of curve changes as we inc. w it goes steeper & steeper which means to give better prob. of class or we can say how sure it is on classification. (prob. - 0/1)

~~means~~ large

Doing little change on x when we have large w leads to class shift. so it leads to overfit ~~on~~ over data



$w = 1$ $\qquad$ $w = 5$ $\qquad$ $w = 100$

Diagram of sigmoid over various w

Ans 1(b)

$$\max_{w_0 \cdots w_d} \prod_{l=1}^{n} P(Y_l \mid X_l, w_0 \cdots w_d) \quad \text{- MLE for logistic Regression}$$

$$\max_{w_0 \cdots w_d} \prod_{l=1}^{n} P(Y_l \mid X_l, w_0 \cdots w_d) \cdot \overbrace{P(w_0 \cdots w_d)}^{\text{prior weight}}$$

MAP estimation

$\mathcal{N}(0, I)$ - Gaussian prior for weight vector (I = Identity Matrix)

$w = [w_0 \cdots w_d]^T$ - log conditional posterior.

$\cancel{L(w)}$

$$L(w) = \log P(w) \prod_{j=1}^{n} P(y^j \mid x^j, w))$$

$$P(w) = \prod_{i=0}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w_i^2}{2}\right)$$

MAP $\quad w^* = \underset{w}{\arg\max}\, L(w) = \underset{w}{\arg\max} \left[\sum_{j=1}^{n} \log(P(y^j / x^j, w)) - \sum_i \frac{w_i^2}{2} \right]$

Gradient ascent

$$w_i^{(t+1)} \leftarrow w^t + \underset{\text{learning rate}}{\eta} \left. \frac{\partial L(w)}{\partial w_i} \right|_t$$

$$\frac{\partial L(w)}{\partial w_i} = \frac{\partial}{\partial w_i} \log p(w) + \underbrace{\frac{\partial}{\partial w_i} \log\left(\prod_{j=1}^{n} P(y^j \mid x^j, w)\right)}$$

$$\frac{\partial}{\partial w_i} \log p(w) = -w$$

After putting value of $\frac{\partial L(w)}{\partial w_i}$

$$w_i^{(t+1)} \leftarrow w_i^{(t)} + \eta \left(-w_i^{(t)} + \sum_j x_i^j \left(y^j - P(Y=1 \mid x^j, w^{(t)})\right)\right)$$

Ans 1 C

$$P(Y = y_K | X) \propto \exp\left(w_{K0} + \sum_{i=1}^{d} w_{Ki} X_i\right) \text{ for } K = 1 \ldots K-1$$

a) <u>model for $P(Y = y_K|X)$ is</u>

$$P(Y = y_K | X) = 1 - \sum_{K=1}^{K-1} P(Y = y_K | X)$$

This is just binary classification

$$P(Y = y_K | X) = \frac{1}{1 + \sum_{K=1}^{K-1} \exp\left(w_{K0} + \sum_{i=1}^{d} w_{Ki} X_i\right)}, \quad K = 1 \ldots K-$$

$$P(Y = y_K | X) = \frac{\exp\left(w_{K0} + \sum_{i=1}^{d} w_{Ki} X_i\right)}{1 + \sum_{K=1}^{K-1} \exp\left(w_{K0} + \sum_{i=1}^{d} w_{Ki} X_i\right)}$$

sum of prob should be one.

<u>Classification Rule</u>

Pick the Rule with highest prob.

$$y = y_{K^*} \quad \text{where } K^* = \arg\max_{K \in \{1 \ldots K\}} P(Y = y_K | X)$$

Ans 1 b d

$$\arg\max_i \exp(a_i) = \arg\max_i a_i$$

max of linear f^n is piecewise linear