① MLE. MAE.

② Weighted LR.

③ MAP.

1. $(x^{(1)}, y^{(1)}), \cdots, (x^{(m)}, y^{(m)})$.

fix

Assumption: $y^{(i)} = \theta^T x^{(i)} + \varepsilon^{(i)}$, $\varepsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$.

$$P_{Y|X}(y^{(i)} | x^{(i)}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right).$$

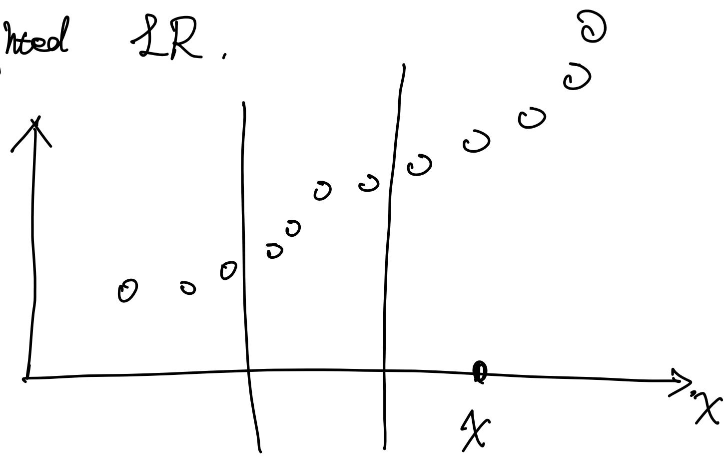
Other Assumption: $\varepsilon^{(i)} \sim \text{Laplace}(\mu, \sigma)$.

$$P_{Y|X}(y^{(i)} | x^{(i)}) = \frac{1}{2\sigma} \exp\left(-\frac{|y^{(i)} - \theta^T x^{(i)} - \mu|}{\sigma}\right). \rightarrow \text{Absolute Value.}$$

$$\sum_{i=1}^{n} \log P_{Y|X}(y^{(i)} | x^{(i)}) = \sum_{i=1}^{n} \log \frac{1}{2\sigma}\left[\exp\left(-\frac{|y^{(i)} - \theta^T x^{(i)} - \mu|}{\sigma}\right)\right].$$

$$= -n \log(2\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n} |y^{(i)} - \theta^T x^{(i)} - \mu|$$

2. Weighted LR.



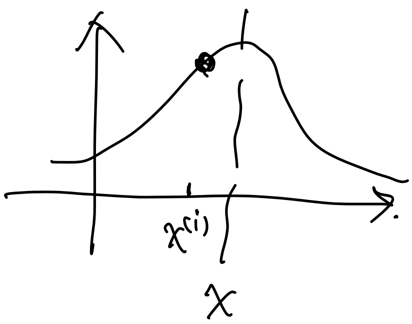
$y = \theta^T x$.

$$J(\theta) = \sum_{i=1}^{m} w^{(i)} (y^{(i)} - \theta^T x^{(i)})^2 \qquad w^{(i)} \in [0, 1].$$

hyperpameter.

One Common choice of $w^{(i)}$:

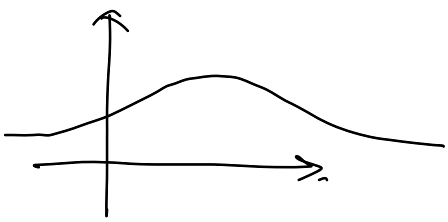
$$w^{(i)} = \exp\left(- \frac{(x^{(i)} - x)^2}{2\sigma^2}\right).$$



i). $x^{(i)} - x$ is small $= 0$. $w^{(i)} \to 1$. $w^{(i)} = 1$.

ii) $x^{(i)} - x$ is large $= \infty$ $w^{(i)} = \exp(-\infty) = 0$.

6.



3. Maximum a postieri (MAP).

X : Data Y: Label, 0~9.

$\boxed{3}$

$P_{Y|X}(0 \mid \boxed{3})$

$P_{Y|X}(1 \mid \boxed{3})$

⋮

$P_{Y|X}(9 \mid \boxed{3})$.

$$P_{Y|X}(y|x) = \frac{P_{XY}(x,y)}{P_X(x)}$$

$$\arg\max_y P_{Y|X}(y|x) = \arg\max_y P_{XY}(x,y).$$

$$= \arg\max_y \underbrace{P_{X|Y}(x|y)}\, P_Y(y)$$

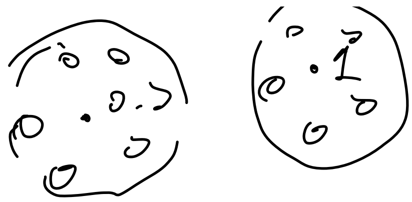
near uniform.

Generative.

$P_{X|Y}(x|y)$. ← model we need to learn.

10000 pics. 10 Label.

1000 $\sim \frac{1}{10}$



WA1:

Bonus.: Truncate SVD.

A: $m \times n$ $m<n$.

rank(A) $= m$.

$$A = [U_1, U_2]\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V \\ 0 \end{bmatrix}.$$