from
$$\hat{y} = f(x; \vec{w})$$

where $\vec{x} = [1 \ x_1 \ x_2 - x_d]$

be the selation between $\vec{y} + x$

we try to minimize the event $\vec{y} + x$
 $\vec{y} = (\vec{y} - \hat{y}^2)^2$ with suspect to $\vec{y} = \vec{y} + x$

best fit for $\vec{y} = f(x; \vec{w})$

tel the optimal $\vec{w} = \vec{y} + x$

I the matrix of inputs be $\vec{y} = \vec{y} + x$
 $\vec{y} = \vec{y} + x$

wix minimizes L(w X^TX is positive definite $\sum_{j=0}^{\infty} \phi_{j} \left(\frac{\ell}{\ell} \right) \cdot \omega_{j}$ ~(E) basis function. being $\phi_{0}(x^{(2)})$ $\phi_{1}(x^{(2)})$ $\phi_{2}(x^{(2)})$ $\phi_{2}(x^{(2)})$ $\phi_{3}(x^{(2)})$ $\phi_{4}(x^{(2)})$ $\phi_{5}(x^{(2)})$ $\phi_{7}(x^{(2)})$ ϕ_{7 $\sum_{\ell=1}^{N} \left(y^{(\ell)} - \sum_{j=0}^{d} \phi_{j}^{(\ell)} . \omega_{j} \right)^{-1}$ $\Rightarrow \sum_{i=1}^{N} \phi_{K}^{(i)} \left(y_{i}^{i} - \sum_{j=0}^{M} \phi_{j}^{(i)} . \omega_{j}^{i} \right)$ $\Rightarrow \sum_{i=1}^{N} (\phi^{T})_{ki} (y - \phi \omega) e = 0$

$$\hat{y}(x, \vec{u}) = u_0 + \sum_{j=1}^{H} u_j \left(2\sigma\left(\frac{2(x-\mu^2)}{5}\right) - 1\right)$$

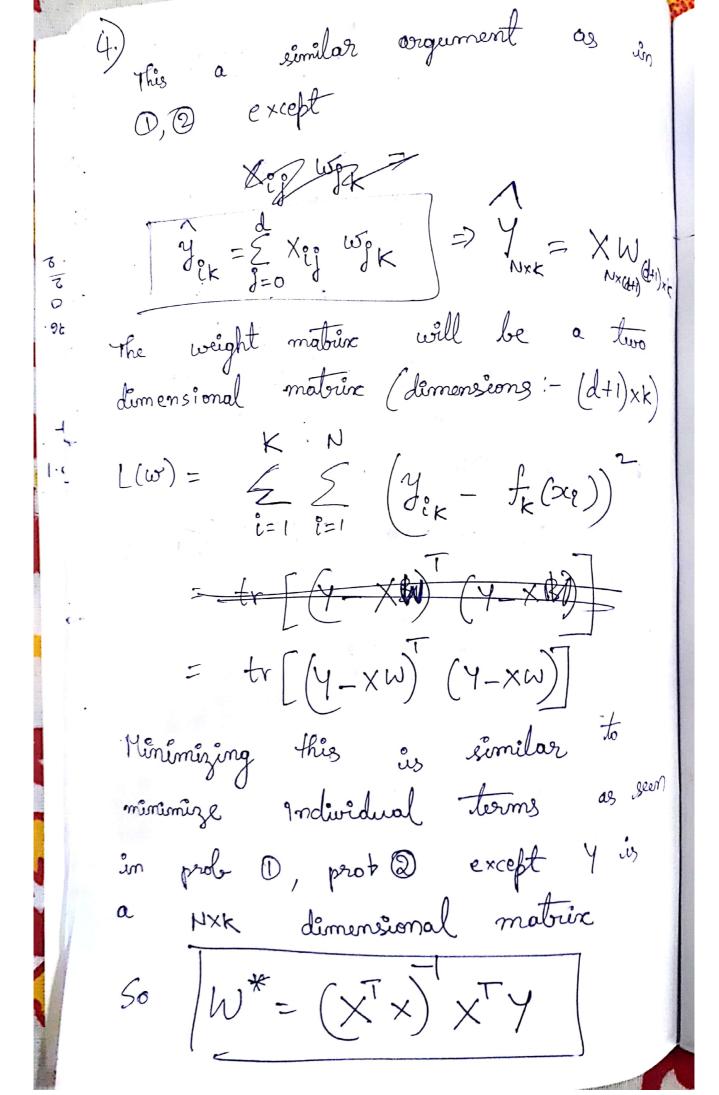
$$= u_0 + \sum_{j=1}^{H} \left(2u_j\right) \sigma\left(\frac{2(x-\mu^2)}{5}\right)$$

$$= \left(u_0 - \sum_{j=1}^{H} u_j\right) + \sum_{j=1}^{H} \left(2u_j\right) \sigma\left(\frac{2(x-\mu^2)}{5}\right)$$

$$= \left(u_0 - \sum_{j=1}^{H} u_j\right) + \sum_{j=1}^{H} \left(2u_j\right) \sigma\left(\frac{2(x-\mu^2)}{5}\right)$$

$$= \left(u_0 - \sum_{j=1}^{H} u_j\right) + \sum_{j=1}^{H} \left(2u_j\right) \sigma\left(\frac{2(x-\mu^2)}{5}\right)$$

$$= \left(u_0 - \sum_{j=1}^{H} u_j\right)$$



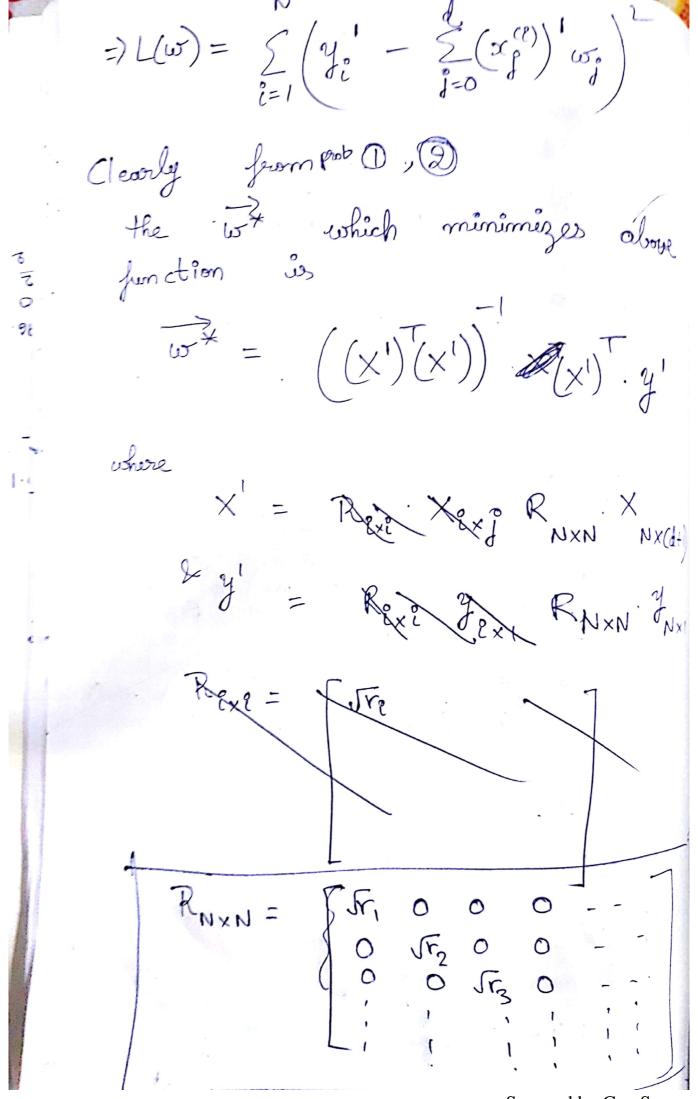
Jeing of limal W

(for bois functions, $X = \phi$)

A is the same module as in proton

(ϕ)

(ϕ) $\frac{5}{2} L(\omega) = \sum_{i=1}^{N} r_i(y^{(i)}) - \sum_{j=0}^{d} x_j^{(i)} \omega_j^{(i)},$ $L(\omega) = \sum_{i=1}^{\infty} \left(\sqrt{x_i} y_i - \sqrt{y_i} x_j^{(c)} \omega_j \right)^{-1}$ y: = Tr: y: 2 (x;))= Tr: x;



$$E(\overrightarrow{w}) = (x^{T} \cdot R \cdot R \cdot x) \times (R^{2})y$$

$$= (x^{T}(R^{2})x) \times (R^{2})y$$

$$= (x^{T}(R^{2})x)$$

use of regularization: > Clearly x x + AI adds a nom-jo term to gettern diagonals of it this makes the Matrix XX+2I invertible for any matrix X Also, this prevents weights from taking extreme values and seducis the high variance of model (better generalization by constraining the morm of weight vector. 7.4 Hore $X_e = X_e + n_e$ is a random variable =) X' is a random vector => X'. w is a random voriable and so gt xi. is y - wo - x. w The dimension of X, N is 1xd &

$$E(\vec{k})^{2}$$

$$= E((\vec{k}, m_{1}, m_{2}, m_{2})^{2})$$

$$= E((\vec{k}, m_{1}, m_{2}, m_{2})^{2})$$

$$= E((\vec{k}, m_{1}, m_{2}, m_{2})^{2})$$

$$= E((\vec{k}, m_{1}, m_{2})^{2})$$

$$= E((\vec{k})^{2} = 0$$

$$= E((\vec{k})^{2})^{2} = E((\vec{k})^{2}) \cdot ((\vec{k})^{2})$$

$$= (\vec{k})^{2} \cdot ((\vec{k})^{2})^{2} = (\vec{k})^{2} \cdot ((\vec{k})^{2})^{2}$$

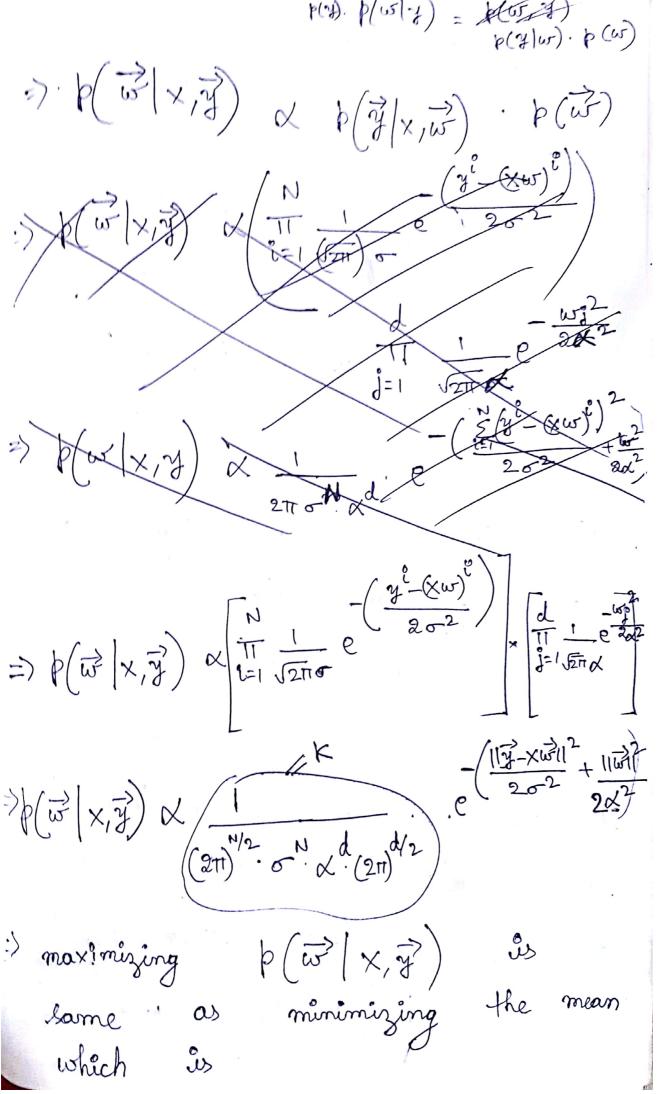
$$= (\vec{k})^{2} \cdot ((\vec{k})^{2})^{2} \cdot ((\vec{k})^{2})^{2} \cdot ((\vec{k})^{2})^{2}$$

$$= (\vec{k})^{2} \cdot ((\vec{k})^{2})^{2} \cdot ((\vec{k})^{2})^{2} \cdot ((\vec{k})^{2})^{2} \cdot ((\vec{k})^{2})^{2}$$

$$= (\vec{k})^{2} \cdot ((\vec{k})^{2})^{2} \cdot (($$

OS Fridge regression loss which is same jundin with Hence proved # (Adding noise to input is same as $(0, \chi^2 I)$

p(w) ~ N(0, x2I) we find to such is that b(w/x,y) is maximized. þ(₹ x, w) ~ N (Xw, σ²) le also y, $\in \overline{y}$ are independent. $\Rightarrow \phi(\vec{y}|x,\vec{\omega})$ $-\left(\begin{array}{c} y^{(l)} - \sum_{j=0}^{d} x^{j}_{j} y^{j} \end{array}\right)$ $= \frac{1}{i=1} \frac{1}{(\sqrt{2\pi})} e$ d - dimension of features No. of train samples. $p(\vec{y}|x,\vec{y}) \sim p(\vec{y}|x,\vec{w}) \cdot p(\vec{w})$ (According to Boye's theorem)



$$L(w) = \frac{\|\vec{y} - x\vec{w}\|^2}{2\sigma^2} + \frac{1}{2\sigma^2}(\|w\|^2)$$

$$= \frac{1}{2\sigma^2}(\|\vec{y} - x\vec{w}\|^2 + \frac{1}{2\sigma^2}(\|w\|^2)$$

$$= \frac{1}{2\sigma^2}(\|w\|^2 + \frac{1}{2\sigma^2}(\|w\|^2))$$

$$= \frac{1}{2\sigma^2}(\|w\|^2 + \frac{1}{2\sigma^2}(\|w\|^2 + \frac{1}{2\sigma^2}(\|w\|^2))$$

$$= \frac{1}{2\sigma^2}(\|w\|^2 + \frac{1}{2\sigma^2}(\|w$$

The same L(w) is also the mode of the distribution (peak value in distribution