Lecture 15-a

*) GPR

*) Regression: (x_i, y_i) $x_i \in R^d$, $y_i \in R^p$

L.S.R. $\min_{W} \| Y - X^T W \|_2^2$

$= \min_{W} \sum_{i=1}^{N} \| y_i - z_i^T W \|_2^2$

$z_i \to i^{th}$ column of X

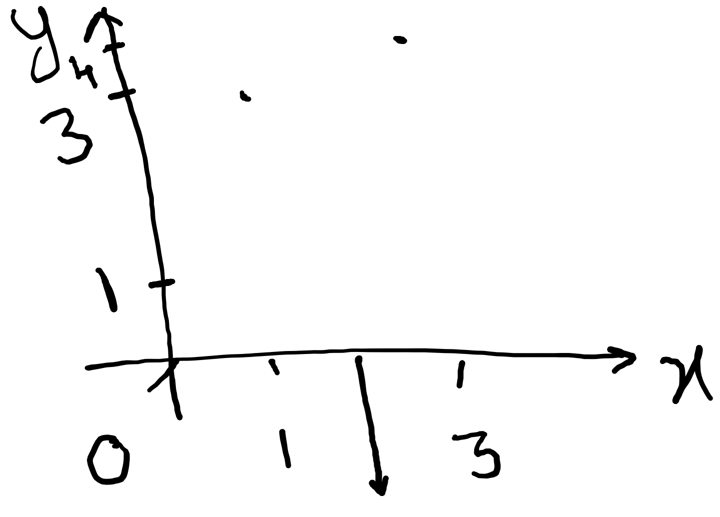
$X \in R^{d \times N}$

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_N \\ 1 & 1 & & 1 \end{bmatrix} \quad W = \begin{bmatrix} w_1 \\ w_0 \end{bmatrix}$$

* GPR

$(0,1)$, $(1,3)$, $(3,4)$

(x, y)



$x_* = 2$, $y_* = ?$

μ_*, σ_*^2

$$\Sigma_{aa} = \begin{bmatrix} \sigma^2 e^{-\|x_1 - x_1\|^2/2l^2} & \sigma^2 e^{-\|x_1 - x_2\|^2/2l^2} & \\ & \sigma^2 & \\ & & \sigma^2 \end{bmatrix}$$

$$\Sigma_{ab} = \begin{bmatrix} K_{1*} \\ K_{2*} \\ K_{3*} \end{bmatrix} = \begin{bmatrix} \sigma^2 e^{-\|x_1 - x_*\|^2/2\theta^2} \\ \vdots \\ \sigma^2 e^{-\|x_3 - x_*\|^2/2\theta^2} \end{bmatrix}$$

$$y_a = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

$$\Sigma_{bb} = \sigma^2$$

*) Regularization → to reduce overfitting
reduces model complexity.

L.S.R.

$$\min_{W} \underbrace{\|Y - X^T W\|_2^2}_{\text{Data fitting}} + \underbrace{\alpha W^T W}_{\text{regularizer}}$$

α is +ve scalar.

$$W^T W = \sum_{i=1}^{d} w_i^2$$

→ $\alpha \to \infty$ → it would focus only on regularizer

$$W = (X^T X + \alpha I)^{-1} X Y$$

Edit $(XX^T + \alpha I)^{-1} XY$

L_1-norm $\rightarrow$ $\|W\|_1 = \sum_{i=1}^{d} |w_i|$

L_2 $\rightarrow$ regularizer $\rightarrow$ Ridge regression / Tikhonov regularization / weight decay

L_1-regularizer $\rightarrow$ Lasso regression.

*) $\|y - X^T w\|_2^2 + \alpha \|w^T w\|_2^2$

$$\min_{w} \; f(w) + g(w)$$

$$\min_{w} \; e^{f(w)} \cdot e^{g(w)}$$

$$\max_{w} \; e^{-[f(w)+g(w)]}$$

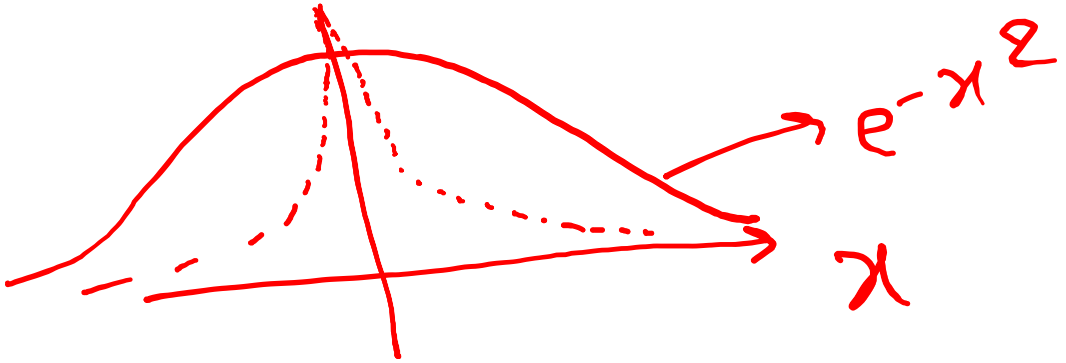
$$\max_{w} \; \ln\left\{ e^{-[f(w)+g(w)]} \right\}$$

$$\max_{w} \; -f(w) - g(w)$$

$$\max_{w} \; e^{-f(w)} \quad e^{-g(w)}$$

$$\max_{w} \; e^{-\|y - X^T w\|_2^2} \quad \underbrace{e^{-\alpha w^T w}}_{\text{Gaussian Prior.}}$$

L_1 reg. $\quad e^{-\alpha \|w\|_1}$



Lasso promotes sparsity.