Exam I : stat (out of 100)

Mean = 80%

Median = 86%

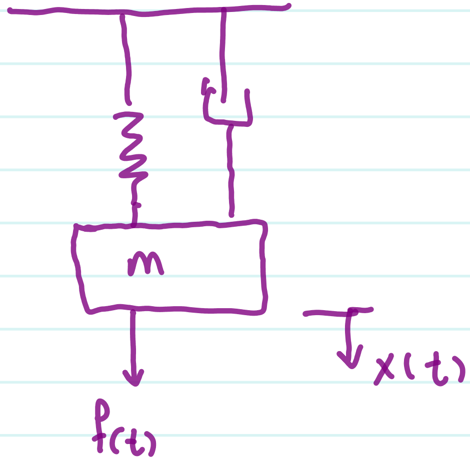
Min = 10%

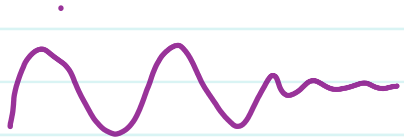
Max = 98%

Std = 18.3%

$\Bigg\}$ HW , Exam I , Exam II , Comp. Final

Top 3



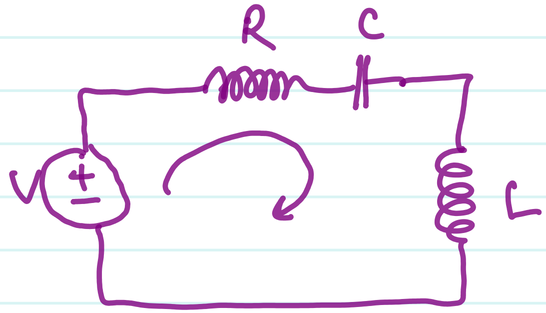


$$F = m\,a = m\,\ddot{x}(t)$$

$$\dot{x} = \frac{dx}{dt} = v(t)$$

$$\ddot{x} = \frac{d\dot{x}}{dt} = a(t)$$

$$V_R + V_C + V_L = V$$



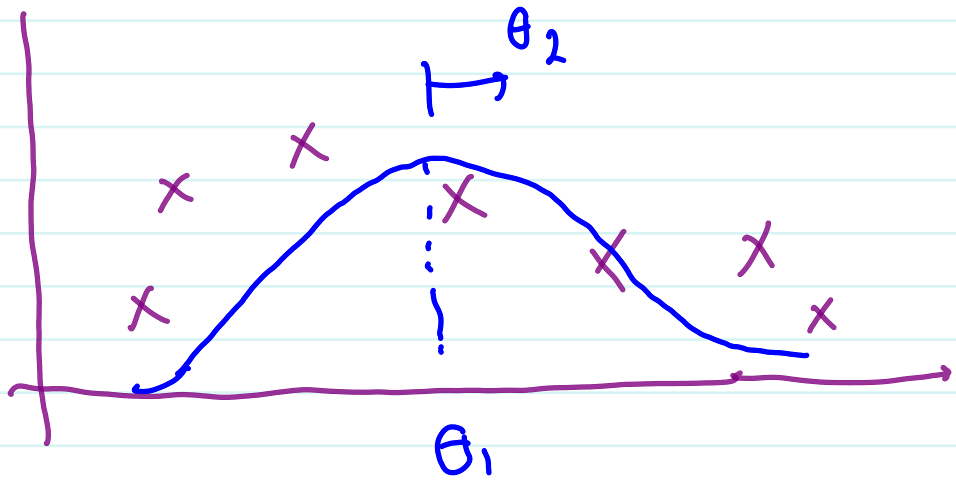
AR (n_a): a_i ?   ARX

$$\underbrace{y(t)}+a_1 y(t-1)+a_2 y(t-2)+\dots+a_{n_a} y(t-n_a)=e(t)$$

$$y(t)+0.5\,y(t-1)=e(t)$$ $e(t)\sim WN(0,\sigma_e^2)$

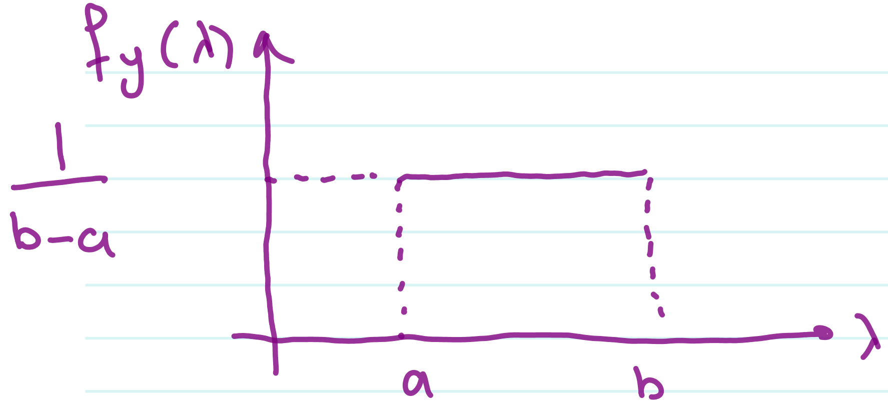
$$\boxed{y(t)=-0.5\,y(t-1)+e(t)}$$

Box-Jenkins Model



θ_2
θ_1

Suppose our data $y_1, \ldots, y_N$ are independently drawn from a uniform distribution $U(a, b)$. Find the MLE of $\hat{a}, \hat{b}$.



$$f_y(\lambda) = \begin{cases} \frac{1}{b-a} & : 0<a<\lambda<b \\ 0 & : \text{Else} \end{cases}$$

$$\theta = \begin{bmatrix} a \\ b \end{bmatrix}$$

y_i are indepedant

$$f_{y_1, \ldots y_N}(\lambda_1, \ldots \lambda_N; \theta) = f_{y_1}(\lambda_1; \theta) * \cdots f_{y_N}(\lambda_N; \theta)$$

$$= \begin{cases} \left(\frac{1}{b-a}\right)^N & : 0<a<\lambda_i<b \\ 0 & : \text{Else} \end{cases}$$

Joint $\longrightarrow$ Likelihood function $\qquad y_i \leftarrow \lambda_i$

$$L(\underset{\sim}{y};\theta) = f_{y_1}(y_1;\theta) \times \cdots f_{y_N}(y_N;\theta)$$

$$= \begin{cases} \left(\frac{1}{b-a}\right)^N : & a < y_i \quad \forall i \in [1, N] \\ & b > y_i \\ \cdot & \end{cases}$$

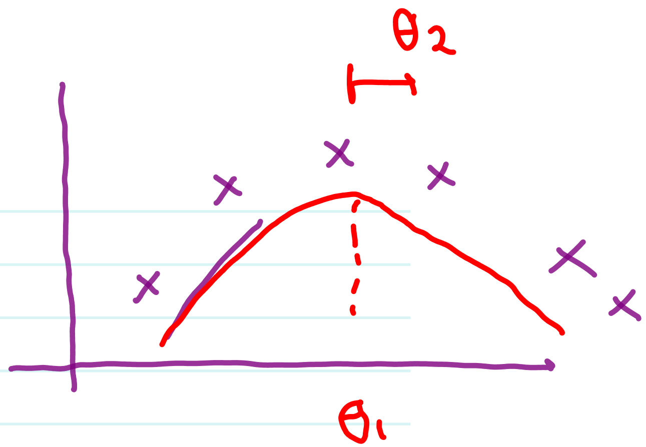
$$a < y_i \quad \forall i \quad \rightarrow \quad \hat{a} = \min\{y_i\}$$

$$b > y_i \quad \forall i \quad \rightarrow \quad \hat{b} = \max\{y_i\}$$

$$\hat{\theta}_{MLE} = \begin{bmatrix} \hat{a}_{MLE} = \min\{y_i\} \\ \hat{b}_{MLE} = \max\{y_i\} \end{bmatrix}$$

Suppose that we have observed random samples $y_1, y_2, \ldots, y_N$ where $y_i \sim N(\theta_1, \theta_2)$. Find the maximum likelihood estimator for θ_1, θ_2

$$f_y(\lambda) = \frac{1}{\theta_2\sqrt{2\pi}} e^{\frac{-(\lambda-\theta_1)^2}{2\theta_2^2}}$$



$y_1, y_2 \ldots y_N = \{y_i\}_{i=1}^{N}$

$\theta = \begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix} = \begin{bmatrix}\mu \\ \sigma\end{bmatrix}$

① $f_{y_1, y_2 \ldots y_N}(\lambda_1, \ldots \lambda_N; \theta) = f_{y_1}(\lambda_1; \underline{\theta}) * \cdots * f_{y_N}(\lambda_N; \theta)$

② $y_i \leftarrow \lambda_i$

$$L(\underline{y}, \theta) = \frac{1}{\theta_2\sqrt{2\pi}} e^{-\frac{(y_1-\theta_1)^2}{2\theta_2^2}} * \cdots * \frac{1}{\theta_2\sqrt{2\pi}} e^{-\frac{(y_N-\theta_1)^2}{2\theta_2^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^N * \left(\frac{1}{\theta_2}\right)^N * e^{-\frac{\sum_{i=1}^{N}(y_i-\theta_1)^2}{2\theta_2^2}}$$

$$\ln(L(\underline{y},\theta)) = \underbrace{-\frac{N}{2}\ln(2\pi)} - \underbrace{N\ln(\theta_2)} - \underbrace{\frac{\sum_{i=1}^{N}(y_i-\theta_1)^2}{2\theta_2^2}}$$

$$e^x \cdot e^y = e^{x+y}$$

$$\boxed{\frac{\partial \ln(L(y,\theta))}{\partial \theta_1}} = 0 - 0 - \frac{(-2)\sum(y_i-\theta_1)}{2\theta_2^2} = 0$$

$$f(x,y) = \frac{x^2}{y^2}$$

$$\frac{\partial f}{\partial y} = \frac{0 - 2y \cdot x^2}{y^4}$$

$$\theta_2 \neq 0 \rightarrow \sum(y_i-\theta_1) = 0 \rightarrow \sum y_i - \sum_{i=1}^{N}\theta_1 = 0$$

$$\sum y_i - N\theta_1 = 0$$

$$\boxed{\hat{\theta}_{1_{MLE}} = \frac{\sum y_i}{N}}$$ Sample mean

$$\frac{\partial(\ln(L(y,\theta)))}{\partial\theta_2} = 0 - \frac{N}{\cancel{\theta_2}} - \left(0 - \frac{\cancel{4}\,\cancel{\theta_2}\sum(y_i-\theta_1)^2}{\cancel{4}\,\cancel{\theta_2^4}\;\theta_2^2}\right) = 0$$

$$N = \frac{\sum(y_i-\theta_1)^2}{\theta_2^2} \rightarrow \theta_2 = \sqrt{\frac{\sum(y_i-\theta_1)^2}{N}}$$

Sample std.

population mean $\Rightarrow$ $\mu_y = E[y] = \int_{-\infty}^{\infty} \lambda f_y(\lambda)\, d\lambda$

$$\boxed{\bar{y} = \frac{\sum y_i}{N}}$$

---

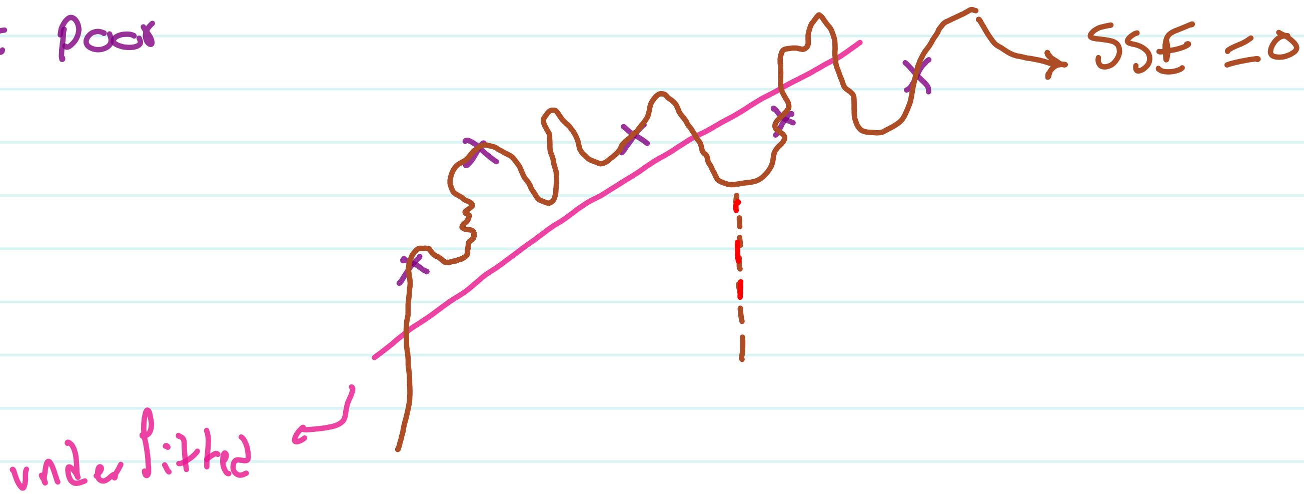
underfitting $\xrightarrow{\text{complexity} \uparrow}$ overfitting

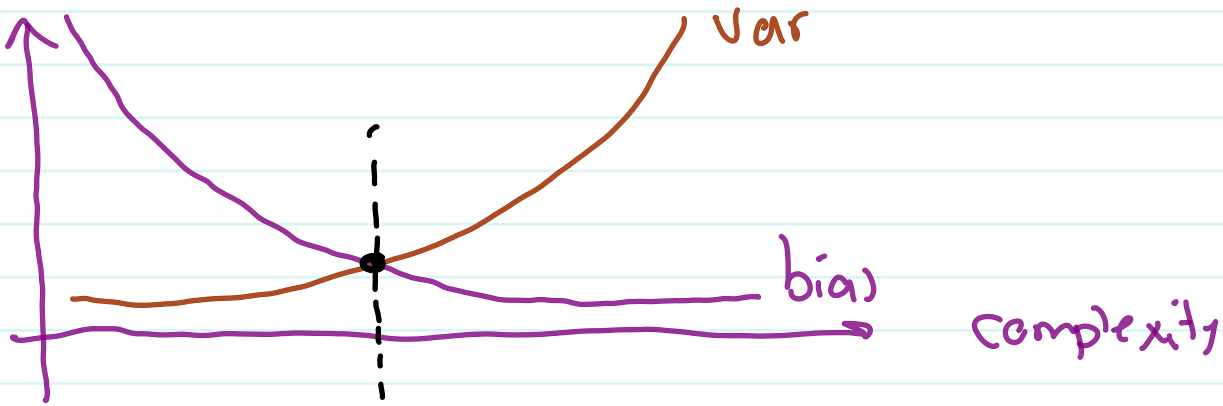
Train = poor

Test = poor

Train : good

Test : poor





$$= X\beta$$

MLR : $$\boxed{\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}$$

Find $\hat{\underline{\beta}}$ $\longrightarrow$ $$SSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} e_i^2 = \underline{e}^T \cdot \underline{e}$$

$$\underline{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$$

$$SSE = (y - X\beta)^T (y - X\beta)$$

$$\frac{\partial SSE}{\partial \beta} = 0 \longrightarrow \boxed{\hat{\beta}_{LES} = (X^T \cdot X)^{-1} X^T \cdot Y}$$

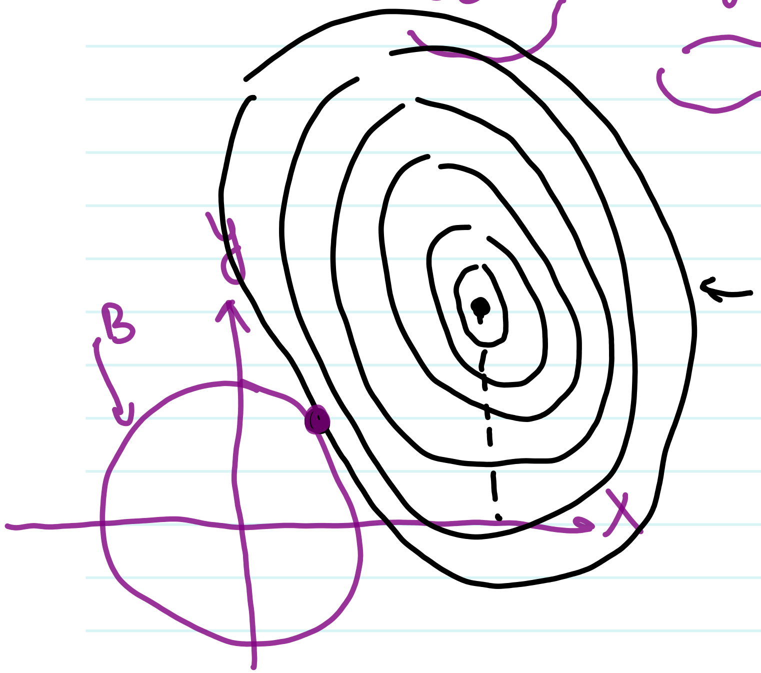
$$\|\underline{e}\|_2^2 = e_1^2 + e_2^2 + \dots + e_N^2 = SSE$$

L_2 - regulirazation:

$$SSE = \|y - XB\|_2^2 + \lambda \|B\|_2^2 \qquad \lambda \in [0, \infty)$$

B

penalty term



proof of L_2 regulirazation :

$$\hat{B}_{MAP} = \underset{B}{\mathrm{Arg\,max}}\, P(B \mid y) = \mathrm{Arg\,max} \frac{\overbrace{P(y \mid B)}^{\text{Likelihood}} \cdot \overbrace{P(B)}^{\text{prior}}}{P(y)}$$

$$\hat{\beta}_{MAP} = \underset{\beta}{\text{Arg max}} \ \ln(P(y|\beta)) + \ln(P(\beta))$$

y_i are observations R.V and has distribution $y_i \sim N(\beta^T x_i, \sigma^2)$

$$f_{y_1,\dots y_N}(\lambda_1,\dots \lambda_N;\beta) = \overbrace{f_{y_1}(\lambda_1;\beta)}\times\dots f_{y_N}(\lambda_N;\beta)$$

$$f_y(\lambda_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\lambda_i - \beta^T x_i)^2}{2\sigma^2}}$$

$\boxed{\text{Likelihood function}}$ $y_i \leftarrow \lambda_i$

$$L(y,\beta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_1-\beta^T x_i)^2}{2\sigma^2}} + \dots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_N-\beta^T x_i)^2}{2\sigma^2}}$$

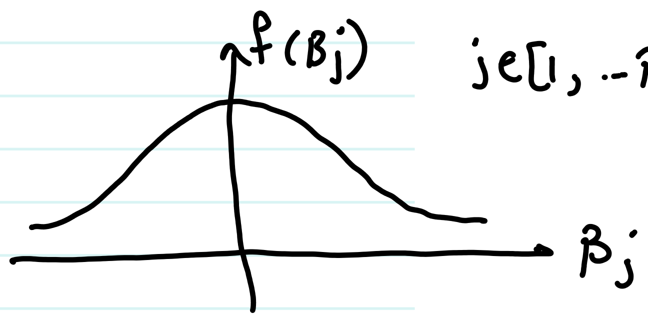
$P(y|\beta)$

$$P(y|\beta) = L(y,\beta) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \cdot \left(\frac{1}{\sigma}\right)^N \cdot e^{-\frac{\sum (y_i - \beta^T x_i)^2}{2\sigma^2}}$$

$$\frac{\partial}{\partial \beta}\left(\ln(L(y,\beta))\right) = \frac{\partial}{\partial \beta}\left(\sum_{i=1}^{N} \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{N} \frac{(y_i - \beta^T x_i)^2}{2\sigma^2}\right)$$

---

what about prior $P(\beta)$; consider a gaussian with $(0, \tau^2)$



$j \in [1, \dots p]$

$$f(\beta_j) = \frac{1}{\sqrt{2\pi}\,\tau} e^{-\frac{\beta_j^2}{2\tau^2}}$$

$$f(\beta_1, \dots \beta_p) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\,\tau} e^{-\frac{\sum_{j=1}^{p} \beta_j^2}{2\tau^2}}$$

$$\ln(f(\underline{\beta})) = \sum_{i=1}^{p} \underbrace{\ln \frac{1}{\sqrt{2\pi}\,\tau}} - \underbrace{\frac{\sum \beta_j^2}{2\tau^2}}$$

So

$$\hat{\beta}_{MAP} = \underset{\beta}{\text{Arg max}} \Big(\overbrace{\ln P(\beta|y)}^{\text{likelihood}} + \overbrace{\ln(P(\beta))}^{\text{prior}} \Big)$$

$$= \underset{\beta}{\text{Arg max}} \left(\sum \underbrace{\ln \frac{1}{\sigma\sqrt{2\pi}}} - \frac{\sum (y_i - \beta^T x_i)^2}{2\sigma^2} + \underbrace{\frac{1}{\sqrt{2\pi}\,\tau}} - \frac{\sum \beta_j^2}{2\tau^2} \right)$$

OR

$$\underset{\beta}{\text{Arg min}} \left(\frac{\|y - X\beta\|_2^2}{2\sigma^2} + \frac{\sum \beta_j^2}{2\tau^2} \right)$$

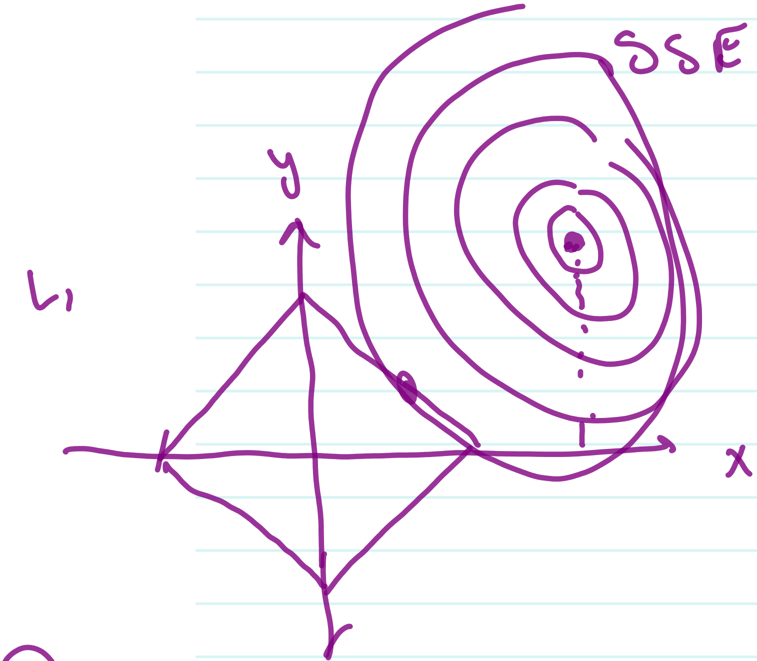
$$\underset{\beta}{\text{Arg min}} \left(\|y - X\beta\|^2 + \overset{\lambda}{\frac{\sigma^2}{\tau^2}} \|\beta\|_2^2 \right)$$

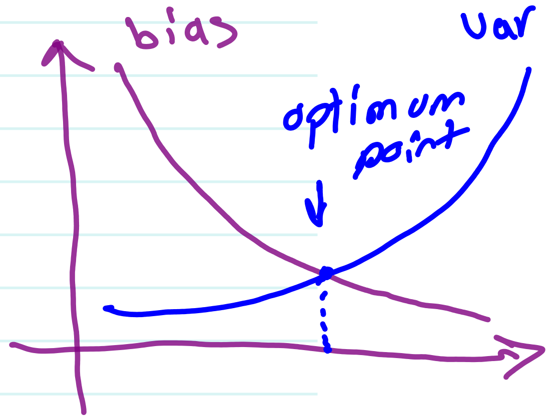
L2 regulirzation

L_1 regularization (B)

Absolut

$$SSE = \|y - XB\|_2^2 + \lambda \|B\|_1$$

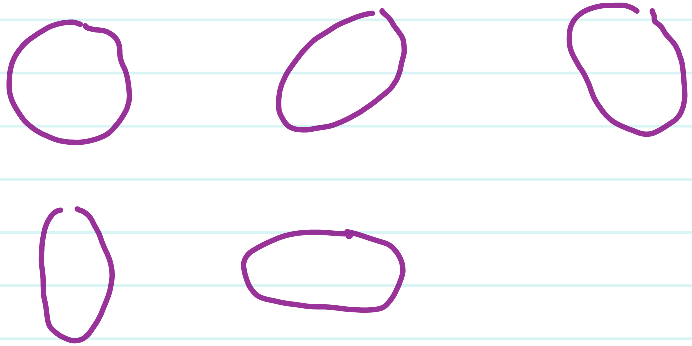




(A)

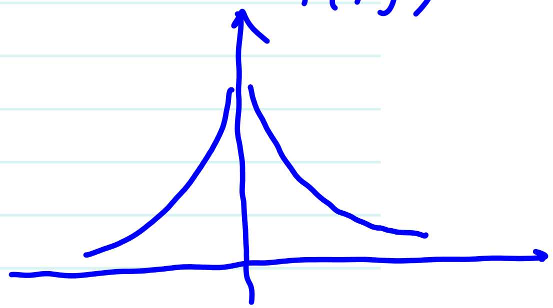
$$SSE = \|y - XB\|_2^2$$

$$B^* = (X^T \cdot X)^{-1} X^T \cdot y$$



$$MSE = Var + bias + \epsilon()$$

L_1 : consider Laplacian distribution



$$\beta_j \sim \frac{1}{2b} e^{-|\beta_j|/b}$$

---

discriminative model

y - red
$\bar{y}$ - blue

features : $x_1, x_2 \dots x_d$

y

$$\underbrace{P(y \mid x_1 \dots x_d) > P(\bar{y} \mid x_1 \dots x_d)}_{\text{red}}$$

$P(y \mid x)$ conditional probability

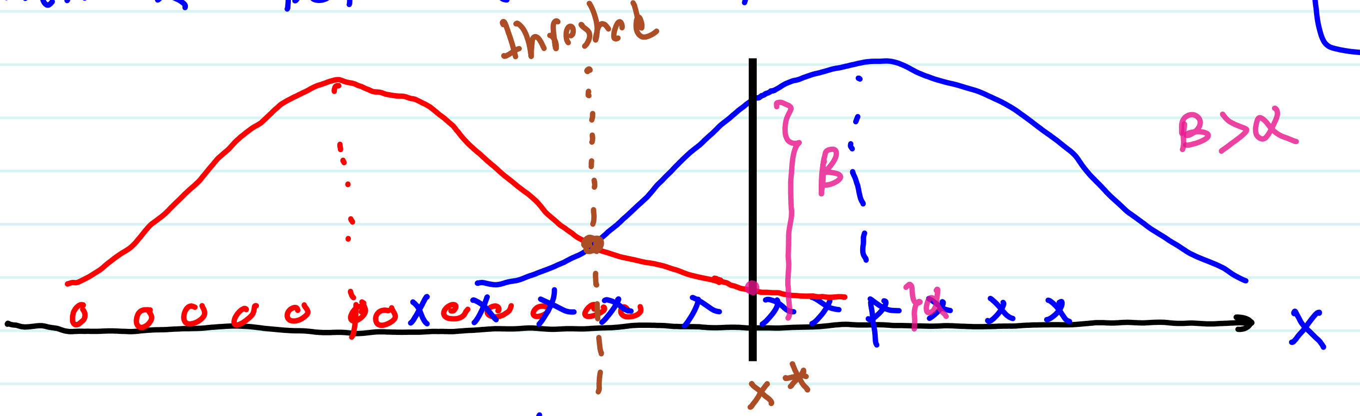
Generative Model

$P(X, y)$ joint probability

given x input we want predict output y

generative model

$P(x, y)$



$P(x^*, red) > P(x^*, blue) \longrightarrow x^* \rightarrow red$

$P(x^*, red) < P(x^*, blue) \longrightarrow x^* \rightarrow blue$

$$\underbrace{P(x, y)} = \underbrace{P(x|y)}_{\text{likelihood}} \cdot \underbrace{P(y)}_{\text{prior}}$$

$P(red) = \frac{1}{2}$

$P(blue) = \frac{1}{2}$

$$P(x|red) \sim N(\mu_{red}, \sigma^2_{red})$$

$$p(X \mid \text{blue}) \sim N(\mu_{\text{blue}}, \sigma^2_{\text{blue}})$$

$$p(x^*, \text{red}) = p(x^* \mid \text{red}) \cdot p(\text{red}) = \alpha \cdot \frac{1}{2}$$

$$p(x^*, \text{blue}) = p(x^* \mid \text{blue}) * p(\text{blue}) = \beta \cdot \frac{1}{2}$$

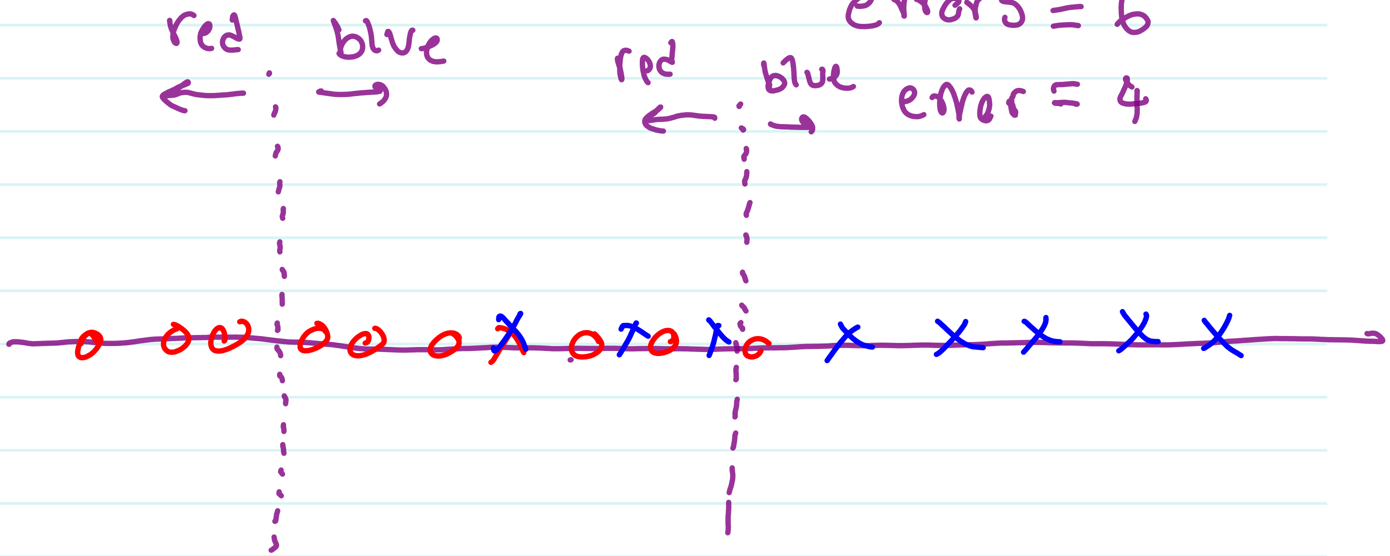
$\beta > \alpha \Rightarrow x^*$ is $\boxed{\text{blue}}$

---

discriminative model

errors = 6

error = 4



generative mode	discriminative model
1 – poor to outliers	1 –
2 – $p(x, y)$	2 – $p(y \mid x)$
3 – Naive bayes	3 – KNN NN SVM
4 – Missing observation is fine	4 – Missing observation is a problem.

Let $(x,y) = \{(1,0), (1,0), (2,0), (2,1)\}$

$P(x,y)$

	$y=0$	$y=1$	
$x=1$	$\frac{1}{2}$	0	$\frac{1}{2}$
$x=2$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
	$\frac{3}{4}$	$\frac{1}{4}$	1

$P(y|x)$

	$y=0$	$y=1$
$x=1$	1	0
$x=2$	$\frac{1}{2}$	$\frac{1}{2}$