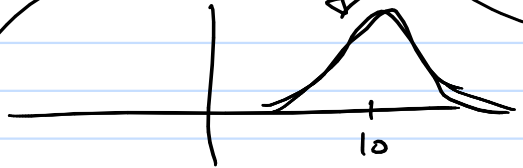
1. MLE and MAP
2. "Everything" about linear classification
3. CV

5-1   4   b

$z_1, \dots z_{30} \sim N(\mu, 1)$

$$\frac{1}{\sqrt{2\pi \cdot 1}} \exp\left(-\frac{(z-\mu)^2}{2}\right)$$



any μ, $P_\mu(Z = z_1) =$

Likelihood of data $= P_\mu(Z = z_1) \cdot P_\mu(Z = z_2) \dots P_\mu(Z = z_{30})$

$$\arg\max_{\mu} \prod_{i=1}^{30} P_{\mu}(Z=z_i) \equiv \arg\max_{\mu} \log\Big(\underbrace{\prod_{i=1}^{30} P_{\mu}(Z=z_i)}_{>0}\Big)$$

$$\log(abc) = \log a + \log b + \log c$$

$$a > b$$

$$\log a > \log b$$

$$10 > 5$$

$$\max_{\mu} \sum_{i=1}^{30} \underbrace{\log P_{\mu}(Z=z_i)}$$

$$\sum_{i=1}^{30} \Big(\underbrace{\cancel{\log \frac{1}{\sqrt{2\pi}}} - \frac{(z_i-\mu)^2}{2}}\Big)$$

$$\arg\max_{\mu} -\sum_{i=1}^{30} \frac{(z_i-\mu)^2}{2} \equiv \underbrace{\arg\min_{\mu} \sum_{i=1}^{30} (z_i-\mu)^2}$$

$$\sum_{i=1}^{30} 2(z_i-\mu) = 0$$

$$\Rightarrow \mu \cdot 30 = \sum_{i=1}^{30} z_i$$

$$\mu = \frac{1}{30} \sum_i z_i$$

OLS

$$\arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta^T x_i)^2 \equiv \arg\max_{\beta} \text{"Likelihood"}$$

$$Y = \beta^T X + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$$

X: Not random

$$Y \sim N(\beta^T X, \sigma^2)$$

Ridge Regression : OLS

MAP : MLE

$$P(\theta \mid \text{evidence}) = \frac{P(\text{evidence} \mid \theta) \cdot P(\theta)}{P(\text{evidence})}$$

No "θ"

$$\min \sum_i (y_i - \beta^T x_i)^2 + \lambda \sum_j \beta_j^2$$

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2 & 0 \\ 0 & \tau^2 \end{bmatrix}\right)$$

$Y \sim N(\beta^T X, \sigma^2)$ $\quad$ $\beta \sim N(0, \tau^2 I)$

λ

LASSO $\quad \lambda \sum_j |\beta_j| \leftarrow$ Laplace dist

② Lin Reg for Classification.

$$G \left\langle \begin{matrix} \text{dog} \\ \text{fox} \\ \text{cat} \end{matrix} \right.$$

$$Y = \underbrace{\begin{bmatrix} \mathbb{1}[G \text{ is dog}] \\ \mathbb{1}[G \text{ is fox}] \\ \mathbb{1}[G \text{ is cat}] \end{bmatrix}}_{3\times 1}$$

$$X_{N\times p} \qquad Y_{N\times K}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$Y_{K\times 1} = \beta_{K\times p} X_{p\times 1}$$

$$Y = K\times 1$$

$$\hat{\beta}^T = \underbrace{(X^T X)^{-1} X^T Y}_{p\times K}$$

$$\hat{G}(x) = \underset{j=1..K}{\operatorname{argmax}} \begin{bmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_K \end{bmatrix} \begin{matrix} -5 \\ -4 \\ 3.1 \end{matrix} \qquad \hat{Y} = \hat{\beta} X^{test}$$

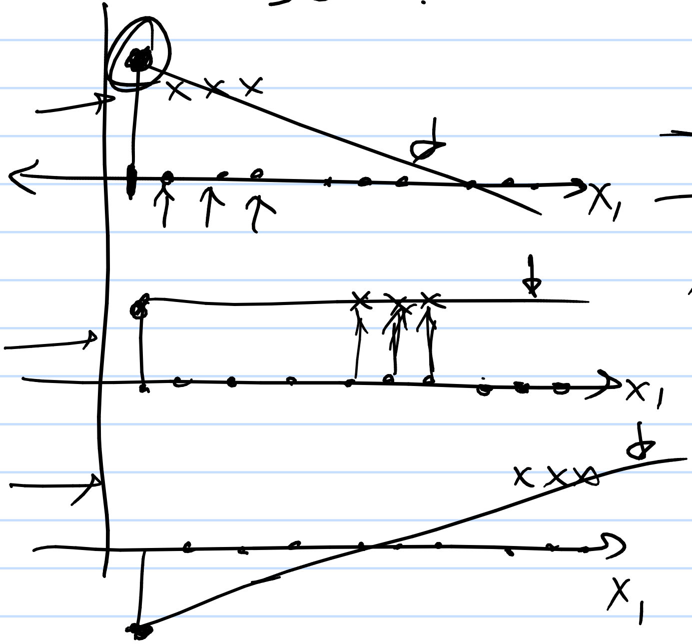
$$Y = \begin{bmatrix} Y_1 = \mathbb{1}[G = dog] \\ \vdots \\ Y_K = \end{bmatrix} \qquad E[\mathbb{1}_{[G=dog]} \mid X = x] = P(G = dog \mid X = x)$$

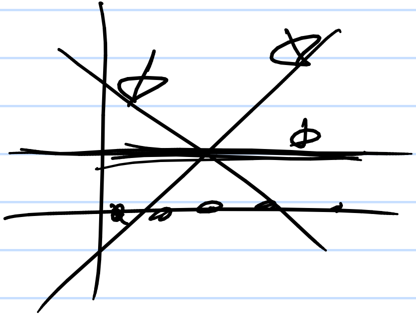
$$\text{★} \quad Pr(G = dog \mid X = x) = E[Y_1 \mid X = x] = \text{linear in } X$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \underbrace{\begin{bmatrix} \quad \\ \quad \\ \quad \end{bmatrix}}_{k \times p} X_{p \times 1} \qquad \boxed{Y_1 = \beta_{1st\,row}^T X}$$

Class masking

3 classes?

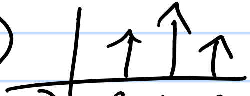




LDA

Generative view

K classes: $G \sim (\pi_1, \ldots, \pi_K)$



$P(G=k)$ $.3 \; .4 \; .3$

g_i

$X \sim P(X=x \mid G=g_i)$ ☆

$$P(G=k \mid X=x) = \frac{P(X=x \mid G=k) \cdot P(G=k)}{P(X=x)}$$

$$P(X=x) = \sum_{k=1}^{K} P(X=x \mid G=k) P(G=k)$$

LDA π_1, π_2, π_3

$$f_k(x) = \boxed{P(X = x \mid G = k)} \quad k = 1, \ldots, K$$

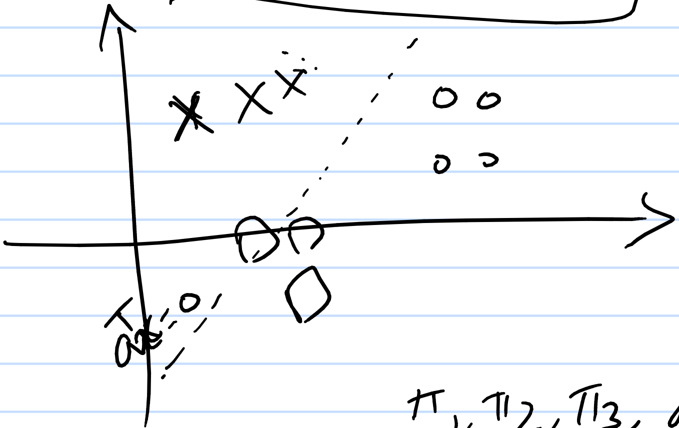
$$= N(\mu_k, \Sigma)$$

$$\log\left(\frac{P(G = l \mid X = x)}{P(G = k \mid X = x)}\right) = \text{linear in } \cancel{\mu_k} \text{ and } \cancel{\mu_l}$$

x

$$\log\left(\frac{P(G=\text{dog} \mid X=x)}{P(G=\text{cat} \mid X=x)}\right) = \log(1) = 0 = \text{linear in } x$$

$$\underbrace{a^T}_{1\times p} x_{p\times 1}$$

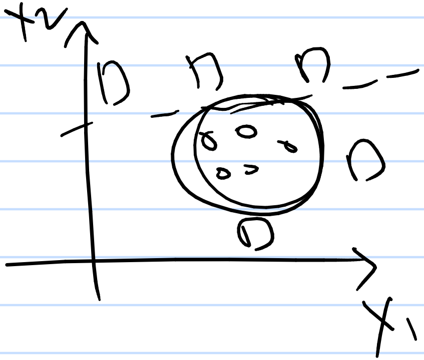


$$\underset{1}{\pi_1}, \underset{1}{\pi_2}, \underset{1}{\pi_3}, \mu_1, \underset{p\times 1}{\mu_2}, \underset{p\times 1}{\mu_3}, \Sigma_{p\times p}$$

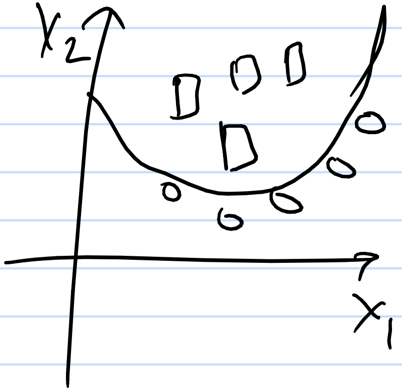
QDA

$G \sim \underline{\pi_1}, \underline{\pi_2}, \underline{\pi_3}$ pmf

$$P(X=x \mid G=\ell) = N(\underline{\underline{\mu_\ell}}, \underline{\underline{\Sigma_\ell}})$$



$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}} \longrightarrow \underbrace{\begin{bmatrix} X_1 \\ X_2 \\ X_1^2 \\ X_2^2 \\ X_1 X_2 \end{bmatrix}} \longleftarrow \text{LDA}$$



$$\underbrace{\{g_i, x_i\}_{i=1}^{N}} = \text{likelihood of data.}$$

$$\hat{\pi}_1 = \frac{\#\text{dogs}}{N} \qquad \hat{\mu}_l = \left(\sum_{i:\text{ is dog}} x_i\right) \times \frac{1}{\#\text{dogs}}$$

MLE

Logistic Reg

$$\log\left(\frac{P(G=l \mid X=x)}{P(G=K \mid X=x)}\right) = \beta_l^T x$$

$l = 1 \ldots K-1$

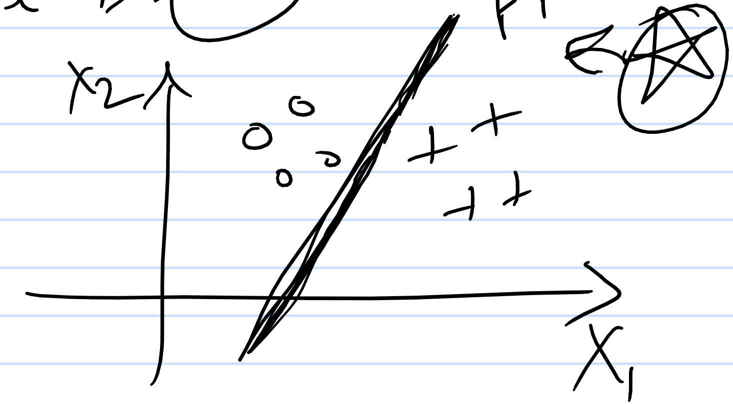
$$\log\left(\frac{P(G=\text{dog} \mid X=x)}{P(G=\text{fox} \mid X=x)}\right) = \beta_{\text{dog}}^T x$$

$$\log\left(\frac{P(G=\text{Cat} \mid X=x)}{P(G=\text{fox} \mid X=x)}\right) = \beta_{\text{Cat}}^T x.$$

$$P(G=l \mid X=x) = \frac{e^{\beta_l^T x}}{1+\sum_{k=1}^{K-1} e^{\beta_k^T x}}$$

$l = 1, \ldots, K-1$

$\beta_1^T x = 0$



$$-\sum_{i=1}^{N} \log P(G=g_i \mid X=x_i)$$

β_1
β_2
$\vdots$
β_{K-1}

max Conditional likelihood

$$\underline{\text{loss}} = -\sum_{i=1}^{N} \left[\log \boxed{P(G=1 \mid X=x_i)}\, y_i + \log\left(1 - P(G=1 \mid X=x_i)\right)(1-y_i) \right]$$

$$y_i = \mathbb{1}[g_i = 1]$$

$$= -\sum_{i=1}^{N} \underbrace{\underline{y_i}^T \log \underline{\hat{P}_i}}$$

Model Selection

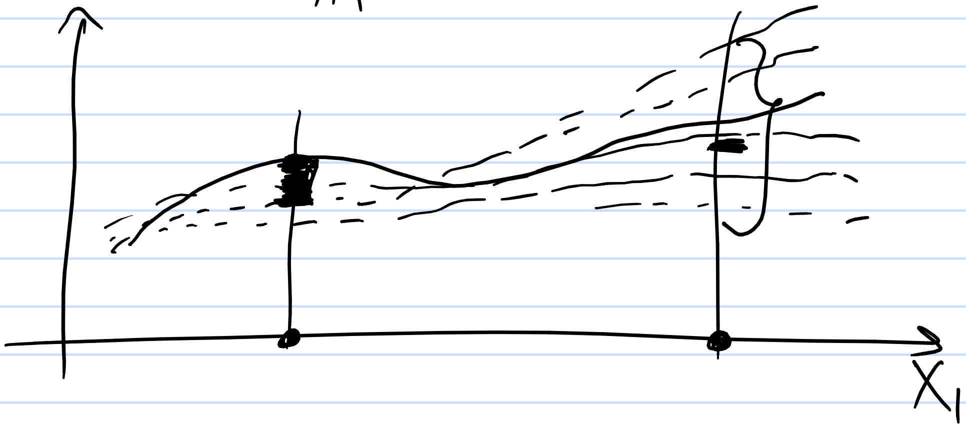
Model Assessment

$$\underline{\underline{Err}} = E_{\mathcal{T}} \, E_{P_{XY}} \left[\text{loss} \left(Y, \hat{f}_{\mathcal{T},\theta}(X) \right) \right]$$

$$\{ \underline{\underline{X_1, Y_1}}, \ldots \underline{\underline{X_N, Y_N}} \}$$

$$= \underline{\underline{\text{irred.}}} + \underline{\underline{\text{Bias}^2}} + \underline{\underline{\text{Var}}}$$

$$(P_{X_1 Y_1} \longrightarrow P_{X_N Y_N}$$



$\underbrace{\{x_i, y_i\}_{i=1}^{N}}_{D} \longrightarrow \hat{f}$

$E_{P_{XY}}\left[\text{loss}\left(Y, \hat{f}(X)\right) \mid D\right] \longleftarrow$ Test error

$E_Z \, E_{P_{XY}}\left[\quad \right] = \text{Err} = \text{Gen}$

---

est Err

CV

Bootstrap

AIC

BIC

MDL

SRM