Bayes' Optimal: $\hat{f}(\underset{\sim}{x}) = \underset{c}{\arg\max}\ p(y=c \mid \underset{\sim}{x})$

SML approach:

1. propose model $\boxed{p_\theta(y=c \mid \underset{\sim}{x})}$
2. need to determine good val. for θ: $\hat{\theta}$

Universal Way: max. likelihood

$\rightarrow \boxed{L(\theta) = \prod_n p_\theta(y_n \mid \underset{\sim}{x}_n)}$

$\rightarrow \ell(\theta) = \log L(\theta) = \sum_n \log p_\theta(y_n \mid \underset{\sim}{x}_n)$

$\rightarrow NLL(\theta) = -\ell(\theta)$

$$\hat{\theta} = \underset{\theta}{\arg\max}\ \underline{\ell(\theta)} = \underset{\theta}{\arg\min}\ \underline{NLL}(\theta)$$

Ex. Lin. Regr. $y_n \mid \underset{\sim}{x}_n \overset{\text{indep}}{\sim} N(\underset{\sim}{x}^T \beta, \sigma^2)$

$$L(\beta) = \prod_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_n - \underset{\sim}{x}_n^T\beta)^2\right)$$

$$\ell(\beta) \propto -\sum_n (y_n - \underset{\sim}{x}_n^T \beta)^2$$

$NLL(\beta) \propto$ Sq. loss btwn y_n and $\underset{\sim}{x}_n^T\beta = \hat{y}_n$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, NLL(\beta) = \underset{\beta}{\operatorname{argmax}}\, \ell(\beta)$$

LDA: $p(y|\underset{\sim}{x}) \propto \underbrace{p(\underset{\sim}{x}|y)}\underbrace{p(y)}$

$$\underset{\sim}{x} | y = c \sim N(\mu_c, \Sigma)$$

$$y \sim Cat(\pi_1, \ldots, \pi_K)$$

$$\theta = (\mu_1, \ldots, \mu_K, \pi_1, \ldots, \pi_K, \Sigma)$$

need to learn good values.

Max ℓ or min NLL

$$\boxed{p(y=c\mid \underset{\sim}{x}) = (2\pi)^{-P/2}\det(\Sigma)^{-1/2}\exp\left(-\tfrac{1}{2}(\underset{\sim}{x}-\mu_c)^T\Sigma^{-1}(\underset{\sim}{x}-\mu_c)\right)\cdot \pi_c}$$

$\boxed{P=1}$

$$p(y=c\mid \underset{\sim}{x}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu_c)^2\right)$$

$$NLL(\theta) \propto \sum_n \left[\frac{1}{2}\log\det(\Sigma) + \frac{1}{2}(\underset{\sim}{x}_n-\mu_{y_n})^T\Sigma^{-1}(\underset{\sim}{x}_n-\mu_{y_n}) - \log\pi_c\right]$$

$$\propto \frac{N}{2}\log\det\Sigma + \frac{1}{2}\sum_n(\underset{\sim}{x}_n-\mu_{y_n})^T\Sigma^{-1}(\underset{\sim}{x}_n-\mu_{y_n}) - \sum_n\log\pi_{y_n}$$

$$\hat{\theta} = \underset{\theta}{\arg\min}\, NLL(\theta)$$

Can show that $\hat{\theta}$ has closed form.

Can show that $\hat{\theta}$ has closed form.

$\rightarrow \hat{\pi}_c = \frac{N_c}{N}$, $N_c = \#$ in class c

% train in class c

$\rightarrow \hat{\mu}_c = \frac{1}{N_c} \sum_{n: y_n = c} \underset{\sim}{x}_n$ mean of $\underset{\sim}{x}$s in class c

$\rightarrow \hat{\Sigma}_c = \frac{1}{N_c} \sum_{n: y_n = c} (\underset{\sim}{x}_n - \hat{\mu}_c)(\underset{\sim}{x}_n - \hat{\mu}_c)^T$

empirical cov. of those in class c

$\rightarrow \hat{\Sigma} = \sum_{c=1}^{K} \hat{\pi}_c \hat{\Sigma}_c$ "pooled" var est.

prediction: $\boxed{\hat{y} = \underset{c}{\operatorname{argmax}}\ \delta_c(\underset{\sim}{x})}$

$\delta_c(\underset{\sim}{x}) = p_{\hat{\theta}}(y = c \mid \underset{\sim}{x})$

$\equiv$ means "up to a transf"

$\equiv p_{\hat{\theta}}(\underset{\sim}{x} \mid y = c)\, p_{\hat{\theta}}(y = c)$

"up to a mono trasf"

$$\equiv p_{\hat{\theta}}(\underset{\sim}{x} \mid y=c)\, p_{\hat{\theta}}(y=c)$$

$$\equiv (2\pi)^{-p/2} \det(\hat{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\underset{\sim}{x}-\hat{\mu}_c)^T \hat{\Sigma}^{-1}(\underset{\sim}{x}-\hat{\mu}_c)\right)\hat{\pi}_c$$

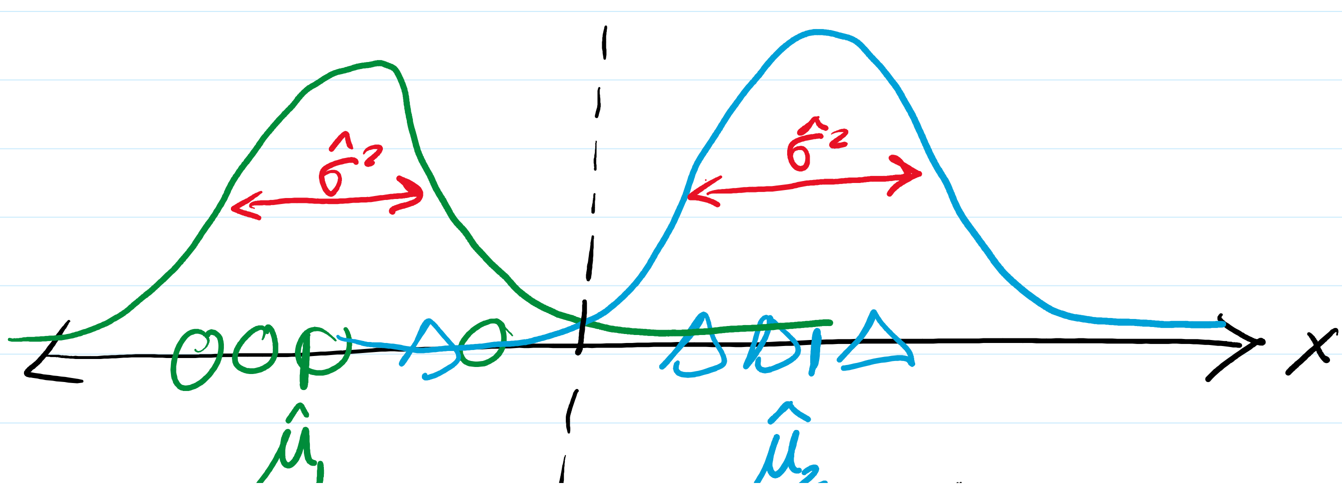
log

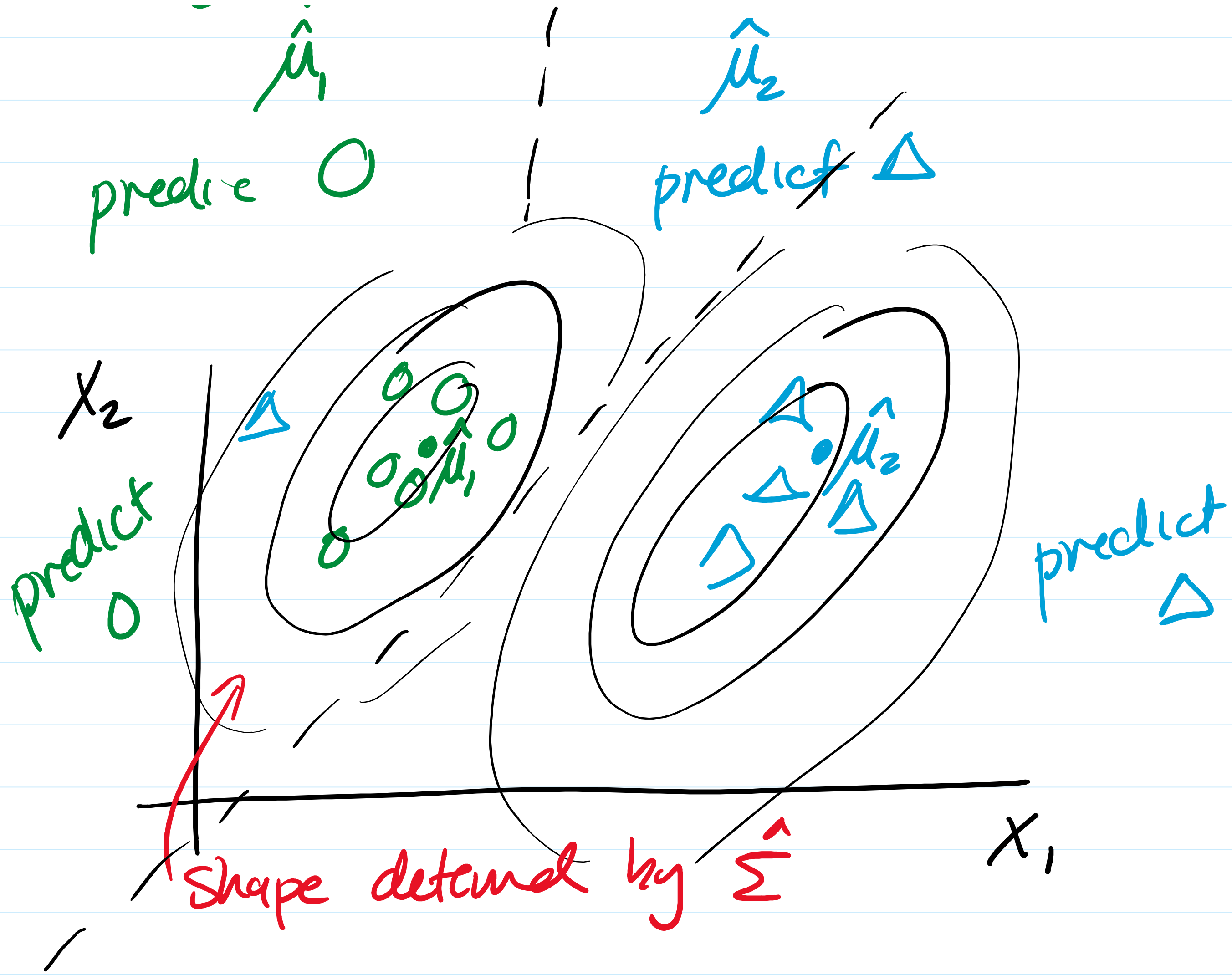
$$\equiv \underbrace{\log \hat{\pi}_c - \frac{1}{2}\hat{\mu}_c^{T}\hat{\Sigma}^{-1}\hat{\mu}_c}_{\hat{\beta}_{0c}} + \underbrace{\hat{\mu}_c \hat{\Sigma}^{-1}}_{\hat{\beta}_c^{T}}\underset{\sim}{x}$$

$$\equiv \hat{\beta}_{0c} + \hat{\beta}_c^{T}\underset{\sim}{x}$$ linear classifier!

1-D case

$$\delta_c(x) = \log \hat{\pi}_c - \frac{1}{2}\hat{\mu}_c^2/\hat{\sigma}^2 + \frac{\hat{\mu}_c}{\hat{\sigma}^2}x$$





QDA: quadratic discrim. analysis

change: $\underset{\sim}{x} \mid y=c \sim N(\mu_c, \Sigma_c)$

← diff cov. mtx per class

⋮

math

⋮

$$\delta_c(\underline{x}) = -\log\det\hat{\Sigma}_c \underbrace{- \frac{1}{2}(\underline{x}-\hat{\mu}_c)^T\hat{\Sigma}_c^{-1}(\underline{x}-\hat{\mu}_c)}_{\text{quadratic}} + \log\hat{\pi}_c$$

<u>\# params</u>: LDA : $(K-1)(P+1) \sim KP$

QDA : $(K-1)\left(\frac{P(P+3)}{2}+1\right) \sim KP^2$

K = # classes

P = # input vars
