Your Data

a data point $\vec{x}_i = \begin{pmatrix} a_i \\ b_i \\ c_i \\ \vdots \end{pmatrix} \Rightarrow \{\vec{x}_i\}$

index of example

$\{(\vec{x}_i, \vec{y}_i)\}$

inputs outputs

random variables

"y"

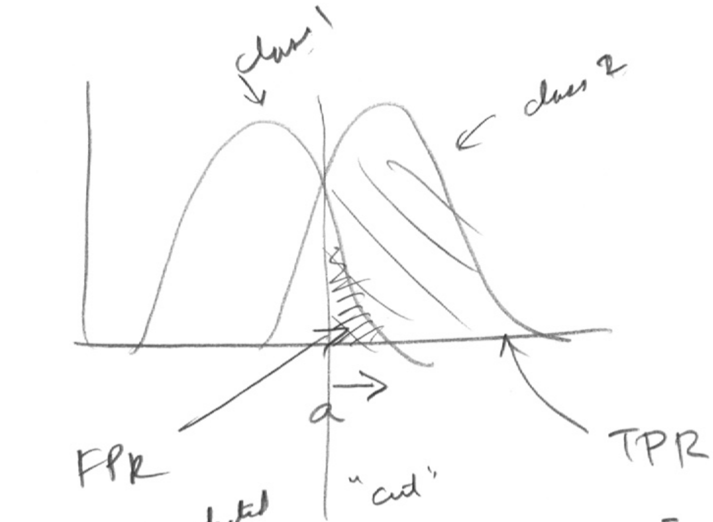
Example index	a	b	c
1	a_1	b_1	c_1
2	a_2	b_2	c_2
3			

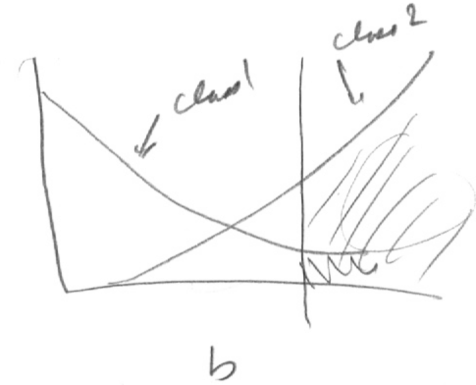
$\Rightarrow$ Continuous / Discrete variables

$\vec{x} \longrightarrow \vec{y} \leftarrow$ Discrete $\Rightarrow$ classification

continuous $\Rightarrow$ Regression

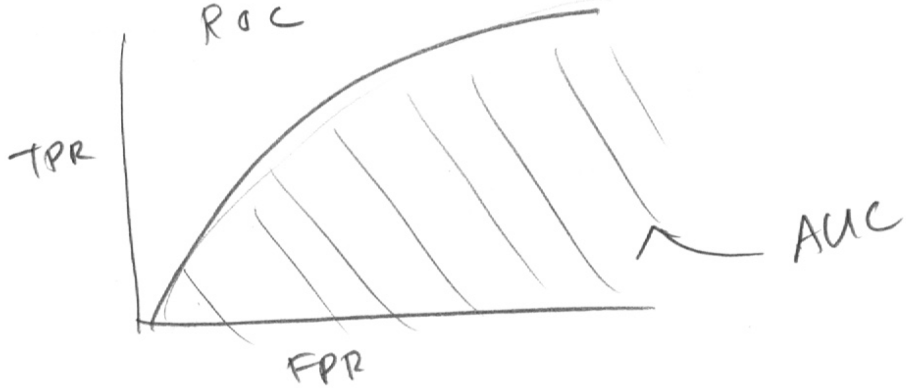
Variable





$$\text{FPR} = \frac{\#\ \text{class 1 selected}}{\text{total class 1}} \qquad \text{TPR} = \frac{\#\ \text{class 2 selected}}{\text{total class 2}}$$

ROC



$\Rightarrow$ Hypothesis testing

Example 1 in 10^6 has measles...

TPR 99%

FPR 1%

$\Rightarrow$ test 10^6 : FP? $10^6 \cdot 10^{-2} = 10^4$

TP? 1

N # of Patients

$\hat{n}$

↑ # of positives

$\bar{b}$ = # of expected false pos.

$\hat{s}$ = # of people w/ measles.

$\bar{b} = FPR \cdot N \cdot r_b$

$\bar{s} = TPR \cdot N \cdot r_s$

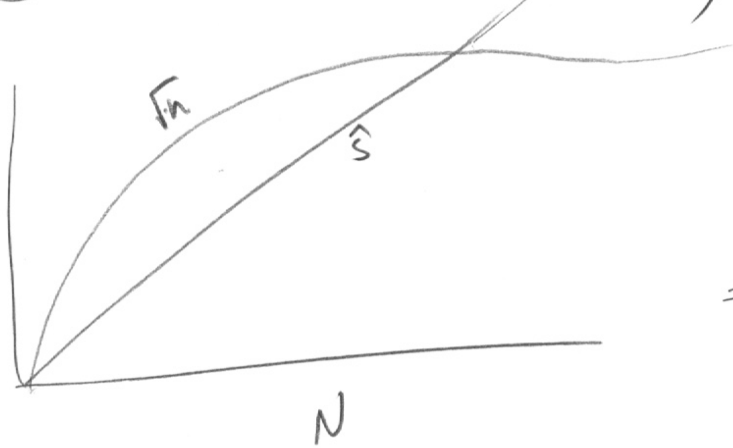
$\hat{s} = \hat{n} - \bar{b}$

$\sigma_{\hat{s}} \approx \sigma_{\hat{n}} = \sqrt{\hat{n}}$

10 ± 100

5 100 ± 100

500 ± 100



$$\Rightarrow \max \frac{\bar{s}}{\underbrace{\sqrt{\bar{s}+\bar{b}}}_{\bar{n}}}$$

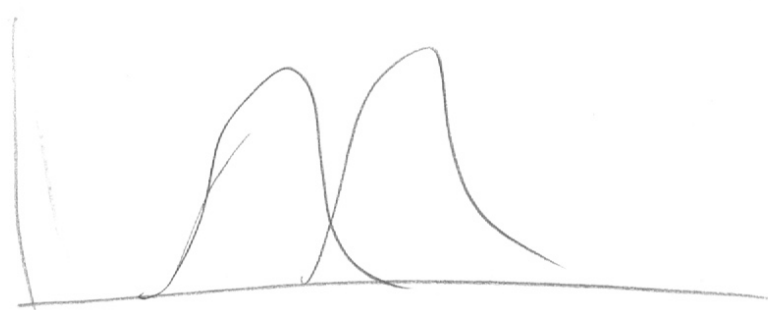
1) "cut" → multi variate

2) Prob Density Estimation (PDE)

$P(\vec{x}_i \mid s)$ $P(\vec{x}_i \mid b)$

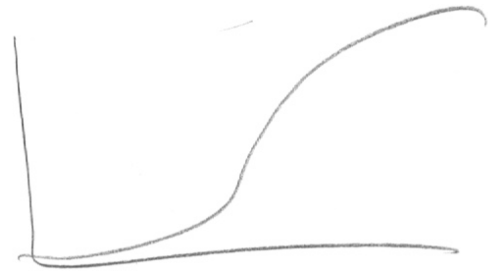
$$\frac{P(\vec{x}_i \mid s)}{P(x_i \mid b)} > c \Leftarrow$$ the best

Neymann Pearson Lemma



curse of dimensionality

3) Test statistic



$$\vec{x}_i \to t_i$$

$$t(\vec{x}_i) = t_i$$

a) $$t(\vec{x}_i) = f(W\vec{x}_i + \vec{b})$$

Linear Comb.

Linear Discriminant Analysis

"Fisher Disc."

$$\vec{W} = (\Sigma_0 + \Sigma_1)^{-1}(\vec{m}_2 - \vec{m}_1)$$

$$F = \vec{W}^T\vec{X} = \vec{w}\cdot\vec{x}$$

$$S = \frac{\vec{w}\cdot\vec{m}_1 - \vec{w}\cdot\vec{m}_2}{\vec{w}^T\Sigma_1 w - \vec{w}\,\Sigma_0 w}$$

$\vec{w}^T(\Sigma_1+\Sigma_0)w$