# Parametric Methods

## Density Estimation

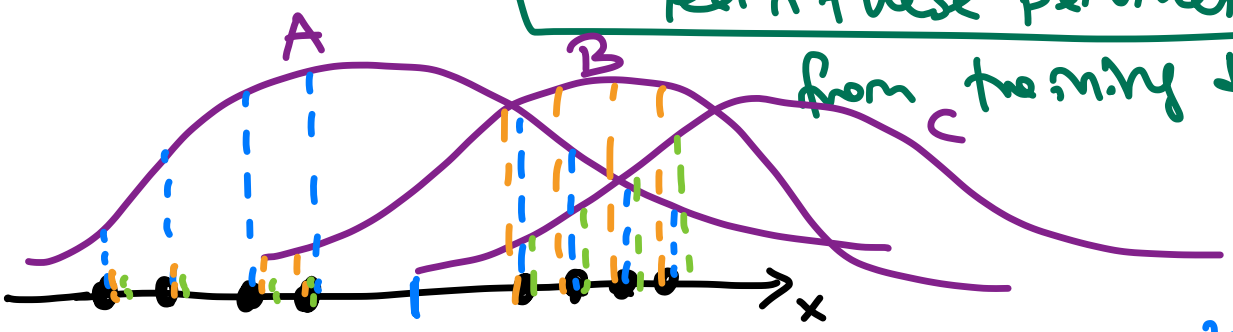
$\mathcal{X} = \{x_i\}_{i=1}^{N}$ N samples / N data points

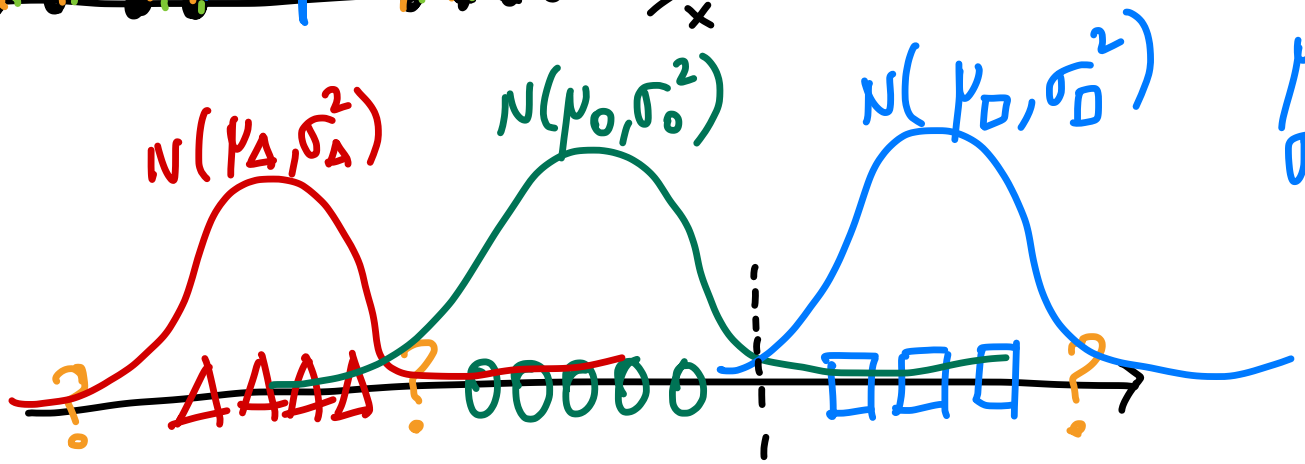
$x_i \sim p(x_i) \quad \forall i$ $\Rightarrow$ probability distribution

$\Downarrow$

parameters (?)

learn these parameters from training data




$x_i \sim N(x_i; \mu, \sigma^2)$

$\mu^*$ : the best $\mu$ parameter

$\sigma^{2*}$ : the best $\sigma^2$ parameter

$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$ $\quad x_i \in \mathbb{R}^1$ $\quad y_i \in \{1, 2, 3\}$

△ ○ □

$p(x \mid y = c) \Rightarrow$ class conditional density

$Pr(y = c) \Rightarrow$ prior distribution

BAYES RULE $\Rightarrow$ $P(B \mid A) = \dfrac{P(A \mid B)\, P(B)}{P(A)}$

$$\underbrace{Pr(\underbrace{y = c}_{B} \mid \underbrace{x}_{A})}_{\text{posterior}} = \frac{p(x \mid y = c)\, Pr(y = c)}{p(x)}$$

$x_{N+1}$ ?

$Pr(y = c \mid x_{N+1})$ $\rightarrow$ $Pr(y = 1 \mid x_{N+1})$, $Pr(y = 2 \mid x_{N+1})$, $Pr(y = 3 \mid x_{N+1})$

pick the maximum one

# MAXIMUM LIKELIHOOD ESTIMATION (MLE)

$\chi = \{x_i\}_{i=1}^{N}$ $\quad x_i \sim p(x_i|\theta) \ \forall i$

↳ unknown parameters of $p$.

$x_i$'s are i.i.d.

↳ identically & independently distributed

Likelihood $\equiv p(x_1, x_2, \ldots, x_N|\theta)$ → full joint

$P(A,B) = P(A)P(B)$

$L(\theta|X) \equiv p(x_1|\theta)\, p(x_2|\theta) \ldots p(x_N|\theta)$

$\equiv \prod_{i=1}^{N} p(x_i|\theta)$

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

$\log(a^b) = b \cdot \log(a)$

log likelihood $\equiv \log\left[\prod_{i=1}^{N} p(x_i|\theta)\right]$

$\equiv \sum_{i=1}^{N} \log\left(p(x_i|\theta)\right)$