AIC & BIC

Mixture Models

Bayesian & conj. prior

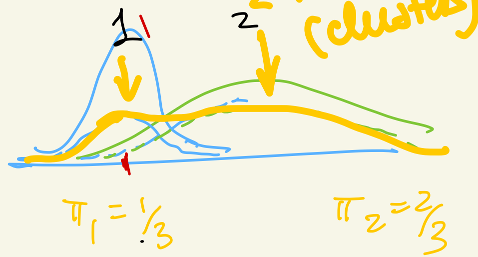
~~Double descent~~

Mixture model

f = weighted sum of K densities

$N(\mu_k, \sigma_k^2)$

$f(x) = \sum \pi_k f_k(x)$



≥ 0 for all x ✓ density

$\int_{-\infty}^{\infty} f(x)\,dx = 1$ ✓

$x \sim f(x)$: sample $k \in 1:K \sim \pi_{1:K}$

sample $x \sim N(\mu_k, \sigma_k^2) = f_k$

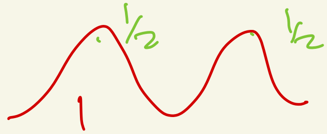
Generative model

parameters:

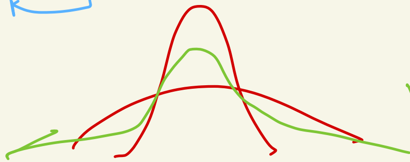
$x \in \mathbb{R}^d$

$\mu_{1:K}$	K	Kd
$\sigma^2_{1:K}$	K	$K\frac{d(d+1)}{2}$
$\pi_{1:K}$	$K-1$	$K-1$
$\sum \pi_k = 1$	$3K-1$	

$\boxed{\Sigma}$



$K=2$



$K=?$

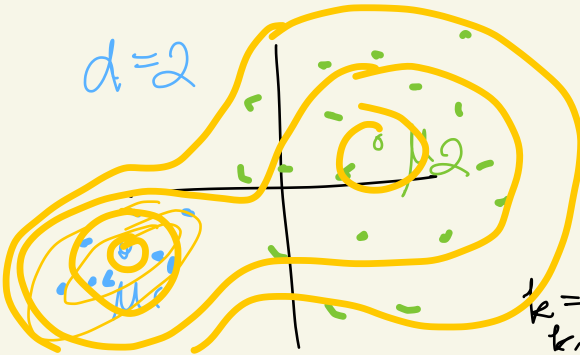
symmetric

Σ

$\in \mathbb{R}^2$

$\mu_{1,2}, \quad \Sigma_{1,2} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$

$d=2$



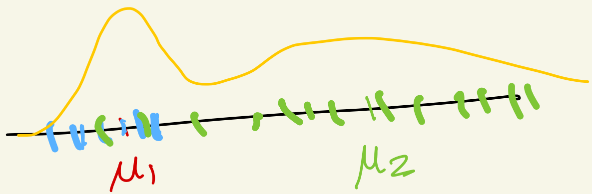
(free) parameters

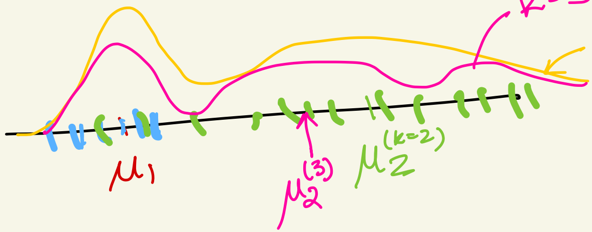
$d^2 - \frac{d(d-1)}{2}$

$d=1$

$k=1 \quad n_1=5$

$k=2 \quad n_2=11$





$f^{(3)}$ true model $k=2$

$\pi^{(3)}_{1,2,3}, \mu^{(3)}_{1,2,3}$ $\Rightarrow$ params: $\underset{\mu}{3} + \underset{\sigma^2}{3} + \underset{\pi}{2}$

estimated model $f^{(2)}$

$\pi^{(2)}_{1,2}, \mu^{(2)}_{1,2}, \Sigma^{(2)}_{1,2}$ # params $= 2+2+1$

$\underset{\mu}{4} + \underset{\sigma^2}{4} + \underset{\pi}{3}$

$K=4 \Rightarrow f^{(4)}$ likelihood

$\ell(\mathcal{D}|f^{(4)}) \geq \ell(\mathcal{D}|f^{(3)}) \geq \ell(\mathcal{D}|f^{(2)})$

$$AIC(f^{(2)}) = \ell(\mathcal{D}|f^{(2)}) - 5$$ ← #params

$$AIC(f^{(3)}) = \ell(\mathcal{D}|f^{(3)}) - (3+3+2)$$

$$AIC(f^{(4)}) = \ell(\mathcal{D}|f^{(4)}) - (\underbrace{4+4+3}_{\text{nr. parameters}})$$

choose max

ML from model class

$$BIC(f^{(4)}) = \ell(\mathcal{D}|f^{(4)}) - \boxed{\frac{\ln n}{2}}(4+4+3)$$

Only parametric (no KDE)

(Parametric)

Model family $\{\text{Normal}\} \xrightarrow[ML]{\text{estimation}} f^{ML} \equiv \mu^{ML}, (\sigma^2)^{ML}$

$\{\text{Logistic}\}$

$\{\text{Poisson}\}$

...

Select model

CV

AIC, BIC

$D \rightarrow$

$\{\text{Mixture } k=2 \text{ Normal}\} \rightarrow f^{ML} \equiv \mu^{ML}_{1,2}, \sigma^2_{1,2}, \pi^{ML}_{1,2}$

$\{$ —"— $k=3$ —"— $\}$

...

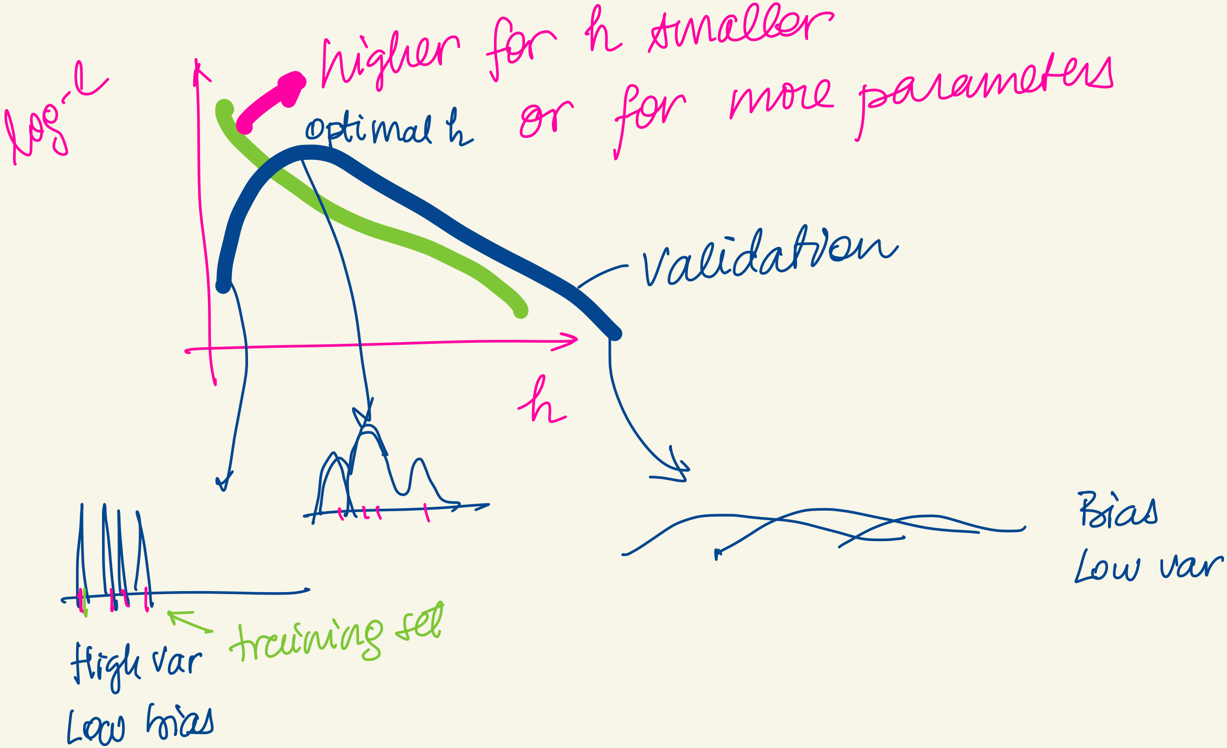
$\{$ —"— $k=3$ Logistic $\}$

...

Nonparametric KDE $\{ f_{h, K(), n} \} \xrightarrow{KDE} f$

Select model

CV



log-L
higher for h smaller
or for more parameters
optimal h
Validation
h
High var
Low bias
training set
Bias
Low var