① Lay of the land

② Lec 2 : Relatives of linear models and k-n-n

K-nn

dimensionality

linear model

P_{XY} P_{GY} EPE

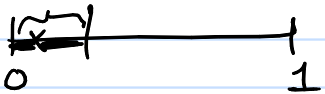
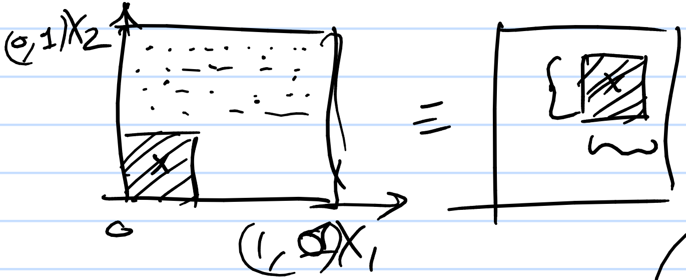
K-nn linear models

SL

SL
non linear

UL

(a) dimensionality & k-nn

	nbhd size	feature range $\equiv$ side length
1. 	$10\% = \left(\frac{1}{10}\right)$	$\frac{1}{10}$
2. 	10%	$\sqrt{\frac{1}{10}}$ =
		$\left(\frac{1}{10}\right)^{1/3}$ =
		$\left(\frac{1}{10}\right)^{1/100} \approx 0.999$

$K(x_i, x^{test})$

(b) k-nn $\longrightarrow$ $\frac{1}{K} \sum_{i=1}^{N} \mathbb{1}\left[x_i \in k\text{-nbhd}(x^{test})\right] \cdot y_i$

linear model $\beta^T X$ eg: $\underline{\beta_1 f_1(X_1)} + \cdots + \beta_p \underline{f_p(X_p)}$

$\beta_1 X_1 + \cdots + \beta_p X_p$ $\qquad \log(X_1) \qquad X_p^2$

$\hat{\beta}$ $\qquad \hat{\beta}$

① Bias Variance trade-off

eg: Linear models

② Cross validation

$P(X=x, Y=y)$ $\qquad$ $P(Y=y \mid X=x)$

linear model $\qquad$ k-nn

more biased $\qquad$ less biased.

$\{x_i, y_i\}_{i=1}^{N}$ = Dataset

$\hat{f}$

P_{XY}

$X \quad Y$

$$EPE = E\left[(Y - f(X))^2\right]$$

$$\hat{f}_\tau \qquad \tau = \{X^i, Y^i\}_{i=1}^{N} \qquad N = 10$$

$$EPE = E_{P_{XY}\,\tau}\left[(Y - \hat{f}_\tau(X))^2\right]$$

Generalization error.

$$\underline{EPE(x_0)} = E_{P_{Y|X=x_0}} \underbrace{E_{\mathcal{Z}}\left[\left(Y-\hat{f}_{\mathcal{Z}}(x_0)\right)^2 \Big| X=x_0\right]}$$

$$\mathrm{Var}(\varepsilon)=\sigma^2$$

$$\underline{Y = f^{true}(X) + \underline{\varepsilon}} \quad (\text{assume})$$

$\varepsilon \sim$ Random var. $E[\varepsilon]=0$

$$EPE(x_0) = \sigma^2 + \underbrace{\left(f^{true}(x_0) - \underbrace{E_{\mathcal{Z}}[\hat{f}_{\mathcal{Z}}(x_0)]}\right)^2} \text{ — } Bias^2$$

$$+ \underbrace{E_{\mathcal{Z}}\left[\left(\underbrace{\hat{f}_{\mathcal{Z}}(x_0)} - E_{\mathcal{Z}}[\hat{f}_{\mathcal{Z}}(x_0)]\right)^2\right]}_{z}$$

$$E_z\left[\left(z - E[z]\right)^2\right]$$

$$\mathrm{Var}_Z\Big(\text{k-nn}(x_0)\Big) \equiv \text{3rd term in previous page.}$$

$$E_Z\left[\boxed{\frac{1}{K}\sum_{l=1}^{K}\left(f^{\text{true}}(\cancel{X}_{(l)}) + \underline{\varepsilon_l}\right)}\right]$$ where $x_1 \ldots x_K$ are the K closest g x_0 $=$

$$Y^i = \underbrace{f^{\text{true}}(X^i) + \varepsilon_i}$$

$$E_Z\left[\hat{f}_Z(x_0)\right] = \frac{1}{K}\sum f^{\text{true}}(X_l)$$

$$\frac{1}{K}\sum_{l=1}^{K} f^{\text{true}}\left(X_{(l)}\right) + \frac{1}{K}\sum_{l=1}^{K}\varepsilon_l$$

$$\hat{f}_Z(x_0) = \frac{1}{K}\left[\sum_{l=1}^{K} f^{true}(X_l)\right] + \varepsilon_l \leftarrow \quad X_l : \text{closest } g \; x_0$$

$$E[X^2] - (E[X])^2$$

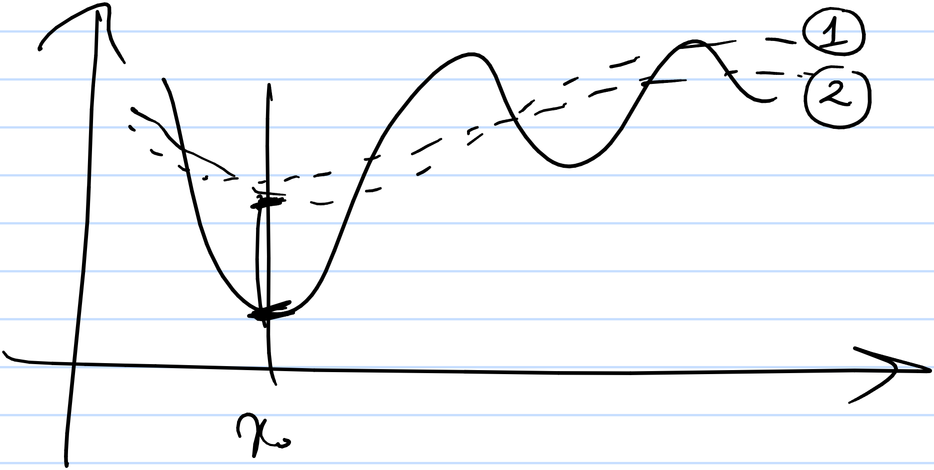
$$E_Z\left[\hat{f}_Z(x_0)\right] = \frac{1}{K}\sum_{l=1}^{K} E_X\left[f^{true}(X_l)\right] \Rightarrow \quad \frac{\sigma^2}{K} \quad \checkmark$$

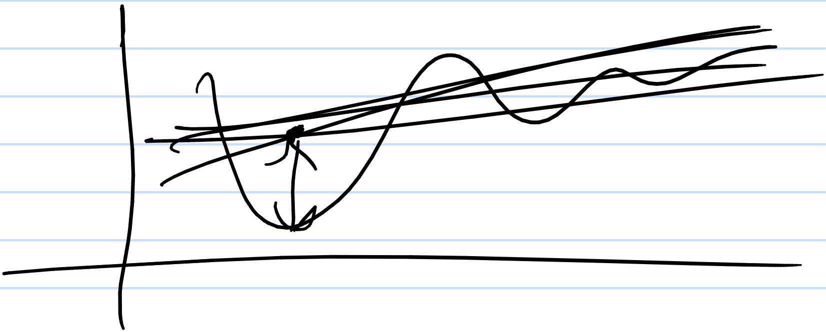
☆ X not random $\quad Y = f^{true}(X) + \varepsilon$

$$\hat{f}_Z(x_0) = \frac{1}{K}(\text{Sum}) + \frac{1}{K}\sum_{l=1}^{K}\varepsilon_l$$

$$E_Z\left[\hat{f}_Z(x_0)\right] = \frac{1}{K}\text{Sum}$$

$$\text{Var}\left(\frac{1}{K}\sum_{l=1}^{K}\varepsilon_l\right) = \frac{1}{K^2}\left(\sigma^2 + \sigma^2 + \cdots + \sigma^2\right)$$


1
2
x_0



1. $\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \theta^T x_i\right)^2$

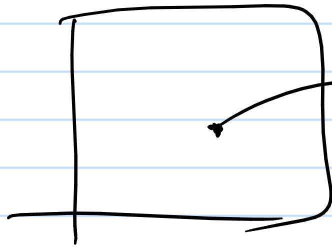
③ $\hat{y} = \sum_{i=1}^{N} y_i$

2. $\sum_{i=1}^{N} \left| y_i - \theta^T x_i \right|$

Yes

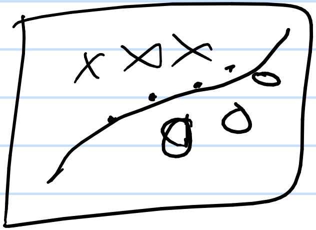
④ $E[Y|X]$

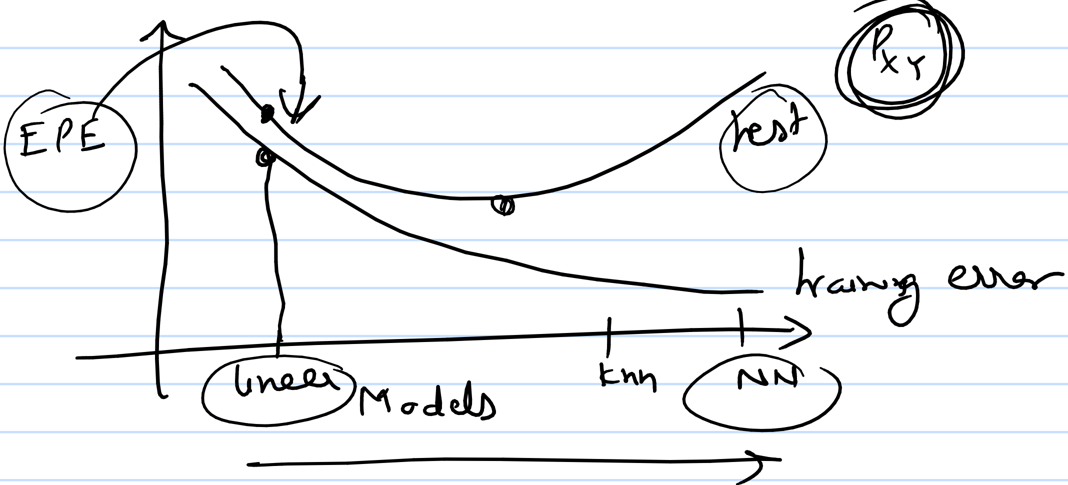
⑤ $\arg\max_{i=1,\dots,K} P(G=i \mid X=x)$

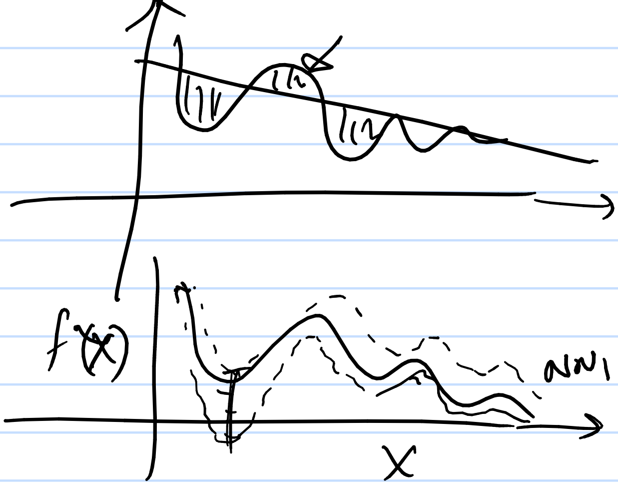


30% orange
70% blue

⑥





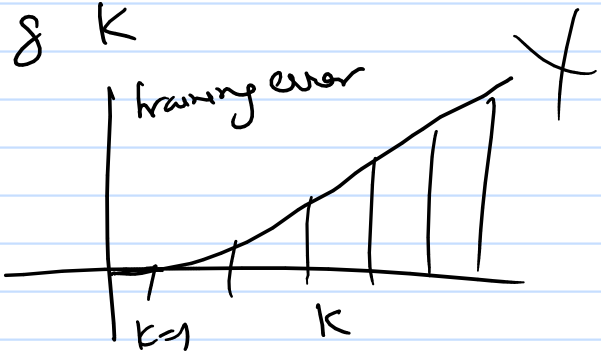


low var.
high bias

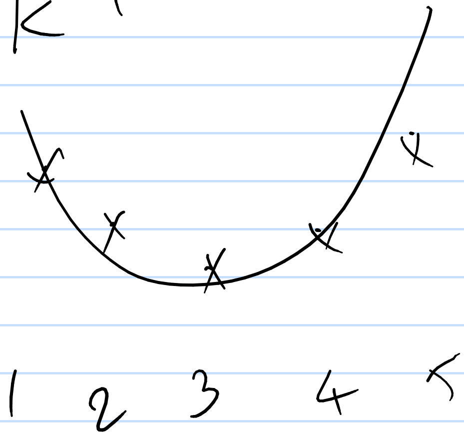
high var
low bias

Choice of K

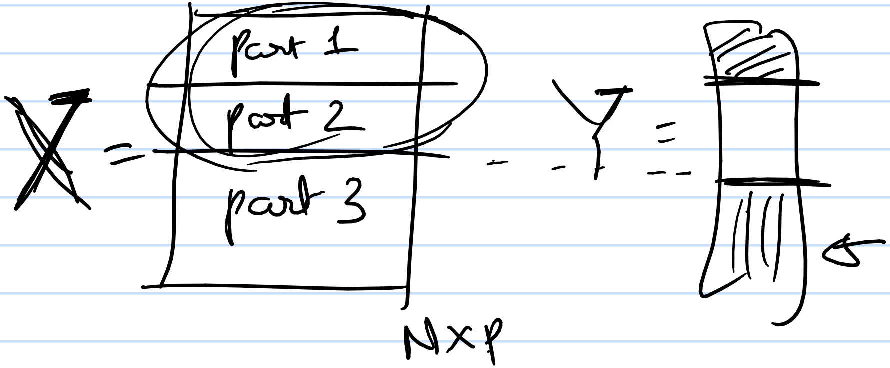
(1)



(2) estimate EPE is via Cross validation



Estimate EPE via Cross validation.



Ⓐ Validate.

pick $k=1$

$k=2$

$\vdots$

$k=15$

$1 \& 2 \Rightarrow 3$ error

$1 \& 3 \Rightarrow 2$ error

$2 \& 3 \Rightarrow 1$ error

RSS

= average the errors

score for my choice of $k=1$

estimate for EPE

$(y_i - f(x_i))^2$

EPE

$E(Y - f(X))^2$