Introducing non-linearity.

Linear = LS, Ridge... = $\hat{f}_{lin}(\vec{x}) = \vec{x}^T \cdot \vec{\beta}^*_{(LS, \text{Ridge}, \ldots)}$ $\qquad \vec{x}^T = (x_1 \ldots x_p)$

Nonlinear = $\hat{f}_{nonlin}(\vec{x}) = \vec{h_m}(\vec{x}) \cdot \vec{\beta}^*_{nonlin}$.

Examples: $\{ h_{m,1}, \ldots h_{m,T} \}$ from $(x_1^{d_1} x_2^{d_2} \ldots x_p^{d_p})$

$\sum d_i < D$

Dictionary

$\binom{p+D}{D}$

Sparsity is desirable.

others: (i) $\log(x_i)$ ...

(ii) $\sin(x_i)$ $\cos(x_i)$ $\sin(kx_i)$

Global vs. local models

$x_1^{(i)} \dots x_p^{(i)} \mid y^{(i)} \qquad i = 1, \dots, N.$

Global models defined by $\vec{\beta}$

Linear models:

$$\hat{y} = \hat{f}_{lin}(\vec{x}^{in}) \qquad \vec{x}^{in} = (x_1 \dots x_p)$$

$$= \vec{x}^{in\,T} \cdot \vec{\beta}^*$$

Nonlinear models:

$$\hat{y} = \hat{f}_{nonlin}(\vec{x}^{in})$$

$$= \vec{h}(\vec{x}^{in})^T \cdot \vec{\beta}_h$$

Alternative is to try and describe the data "locally" (or piece-wise)

We aim not for a model $\hat{f}$ but for a model that is different depending on $\vec{x}^{in}$

$$\left\{\hat{f}_{j(\vec{x}_{in})}\right\} \rightarrow \hat{y} = \hat{f}_{j(\vec{x}^{in})}(\vec{x}^{in})$$

For continuous variables: $\vec{x}^{(i)} \in \mathbb{R}^p$, $y \in \mathbb{R}$

$i = 1, \dots, N$

Introduce a metric in the space of inputs:

$$\| \vec{x}^{(i)} - \vec{x}^{(j)} \|$$

Distance is a choice

Given an input $\vec{x}^{in}$:

① compute all distances between $\vec{x}^{in}$ and the samples.

$$\| \vec{x}^{(in)} - \vec{x}^{(i)} \| \quad \forall\, i = 1, \dots, N$$

② Find the K-nearest neighbours which define the ~~neighbo~~ neighbourhood

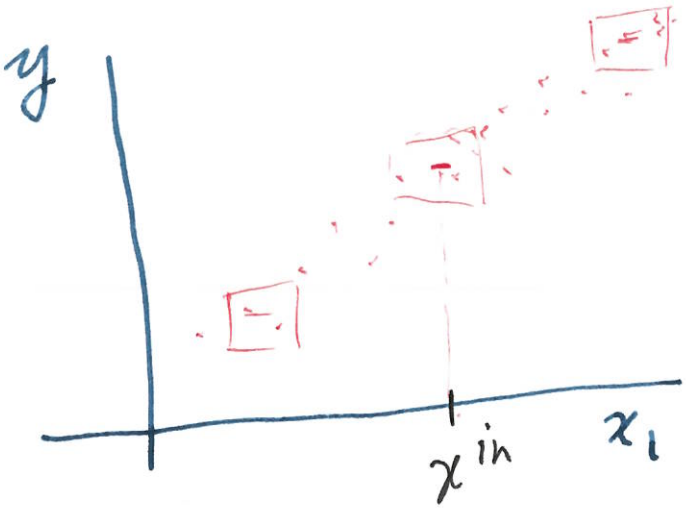
$$\mathcal{N}_K(\vec{x}^{in})$$

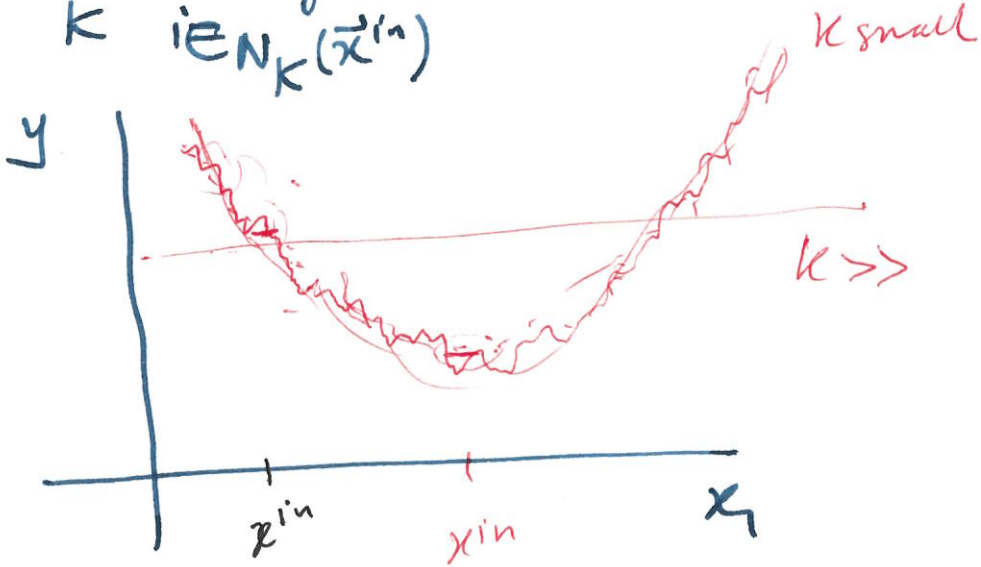
K is a choice

K-NN = K nearest neighbours

(3) Simplest choice for predictor:

$$\hat{f}(\vec{x}^{in}) = \frac{1}{k} \sum_{i \in N_k(\vec{x}^{in})} y^{(i)}$$





The outcome depend on our choice:

(1) Distance and normalization matter a lot.

$x_1 \dots x_p$

$1 < x_1 < 10^6$ $\quad$ $10^{-10} < x_p < 10^{-5}$

(2) It works better for small, relevant p.

(3) K matters a lot.

↳ allows to scan the bias-variance tradeoff.