

- K nearest neighbours
- curse of dimensionality

Until now, we have estimate $f(x)$, from $y = f(x) + \varepsilon$, through linear regression models, in a parametric form.

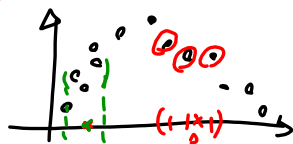
$$\hat{y}_i = f(x; \hat{\beta}) \quad \beta_0 \text{ and } \beta_1$$

- we have decided a specific class of functions $f(x; \beta)$
a.g., lines $\beta_0 + \beta_1 x$

- selected the best instance, $\hat{\beta}_0$ and $\hat{\beta}_1$

ALTERNATIVE: use a non-parametric approach, the function $f(x)$ is approximated only using the data directly

K - nearest neighbours



$\hat{y} = \frac{1}{K} \sum_{i \in N_K(x_0)} y_i$

where $N_K(x_0)$ contains the indexes of the K closest observations to x_0 .

We are modelling $f(x)$ by local constants based on the K closest observations

BASIC IDEA: similar inputs have similar outputs

The method relies on the tuning parameter K , that controls the model complexity

→ larger K , more observations are involved in the estimate, less complex the model. At the extreme, when $K=n$, we estimate the whole function with the mean → null model

→ the smaller the K , the more complex the model. At the extreme, $K=1$, complete interpolation

K is inversely related to the complexity
~~smaller~~ larger K , higher variance, lower bias
~~larger~~ smaller K , lower variance, higher bias



prediction error = $\sigma^2 + \text{bias}^2 + \text{variance}$

$\overline{\text{EPE}}$

in the case of KNN $\sigma^2 + \left(f(x) - \frac{1}{K} \sum_{i \in \mathcal{N}_K(x)} y_i \right)^2 + \frac{\sigma^2}{K}$

$$\text{Var} \left(\frac{1}{K} \sum_{i \in \mathcal{N}_K(x)} y_i \right) = \frac{1}{K^2} \sum_{i \in \mathcal{N}_K(x)} \text{Var}(y_i)$$

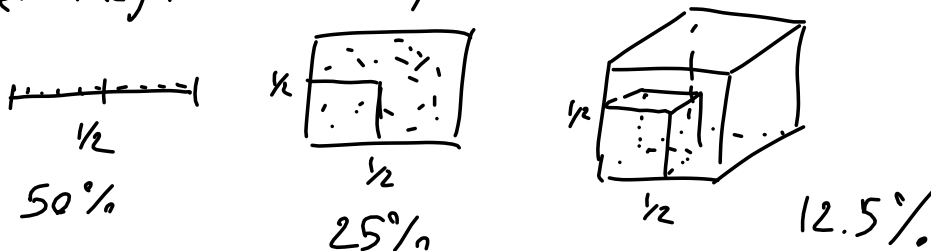
$$= \frac{1}{K^2} K \sigma^2 = \frac{\sigma^2}{K}$$

larger K
→ smaller variance

larger K , more distant points involved in the estimation, less precise, higher bias

K-NN, as a local method, works well when there are many observations. When the number of dimensions (of x) increases, the observations tend to be farther from each other and the K-NN estimate poorer
⇒ CURSE OF DIMENSIONALITY

Example: imagine X uniformly distributed on sample space



$$d(n, p) = \left(1 - \frac{1}{2} \frac{1/n}{d} \right)^{1/p}$$

$n \uparrow$ $d \downarrow$

$\frac{1}{p}$ $\frac{1}{d}$

In order to compensate for the increased distances, we need an increased number of observations to keep the local estimate good enough, the growth is on the order of n^p

Eg.: if we want something similar to what we get with 100 observations in one dimension, we need:

100⁵ with dimension 5 : ten billions

100¹⁰ " " 10