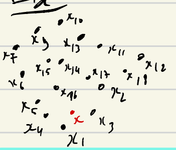
K-nearest Neighbor Classifier

K-nearest neighbor:

$(X, Y) \sim D$

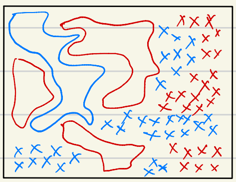
$S_{Tr} = \{(x_i, y_i)\}_{i=1}^{N}$ iid $\sim D\}$

Ex:



$nbh_{S_{Tr}, 4}(x) = \{x_1, x_3, x_4, x_{16}\}$

why is it meaningful?



- meaningful when there is spatial correlation
- Implicitly learn very complex decision boundaries in low dimension!!

Nearest neighbor functions:

$nbh_{S_{Tr}, k}(x) = \{$ set of k elements of S_{Tr} the closest to $x\}$

$\mathcal{X} \longrightarrow \mathbb{N}$

- **Regression**: $y \in \mathbb{R}$

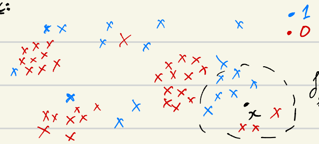
$f_{S_{Tr}, k}(x) = \frac{1}{k} \sum_{i \in nbh_{S_{Tr}, k}(x)} y_i$

$f(x) = \frac{1}{4}(y_1 + y_2 + y_3 + y_5)$

- **Classification**:

$y \in \{0, 1\}$ (Binary classification) $\quad f_{S_{Tr}, k}(x) = \underset{i \in nbh_{S_{Tr}, k}(x)}{\text{majority}} (y_i)$

Ex:

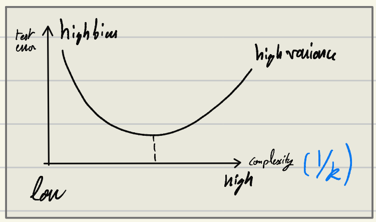


$f_{S_{Tr}, 9}(x) = 1$

- take k odd to avoid tie
- No training method
- Correct drawbacks: add weights (closer better) (weight vote)

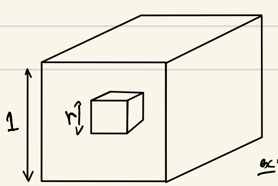
- **Bias variance in function of K.**

- small K, K=1 ⟶ complex model: low bias, high variance
- large K, K = # samples ⟶ simple model (high bias) (low variance)
- complexity scales as $1/k$ ⟶ complexity ↗ when k ↘



Curse of dimensionality

$\mathcal{X} = [0, 1]^d$



- What is the chance that an uniform point in the large box is in the small box

⟶ r^d (very small)

- $(1 - r^d)^N$: none of the samples is inside the small box.

ex: if $N = 500$, $d = 10$ then $r = 0{,}52$ to find a neighbor

Comparing nearest neighbor classifier with the optimal Bayes classifier.

• **Analysis:** $(x,y) \sim D$; $y \in \{0,1\}$ $\quad \mathcal{X} = [0,1]^d$, $x \in \mathcal{X}$

• $\mathcal{L}(f) = \mathbb{P}(y \neq f(x))$

Bayes Predictor:

$$\eta(x) = \mathbb{P}(y=1 \mid x) = 1 - \mathbb{P}(y=0 \mid x)$$

if I know η, how can I predict when given x_0

if $\eta(x_0) \geq 1/2 \longrightarrow 1$

if $\eta(x_0) < 1/2 \longrightarrow 0$

• $f^*(x) = \mathbb{1}_{\eta(x) \geq 1/2}$

$$\mathcal{L}(f^*) = \mathbb{P}(f^*(x) \neq y) = \mathbb{E}_{x \sim D_x}\left[\min\{\eta(x), 1-\eta(x)\}\right]$$ (Bayes loss, (MAP))

• **Claim:**

$$\mathbb{E}_{S_{Train}}\left[\mathcal{L}(f_{S_{Train},1})\right] \leq 2\,\mathcal{L}(f^*) + \text{geometric term}$$ (closest neighbor)

. We assume that $\exists c \geq 0, \forall x, x' \in \mathcal{X}, \; |\eta(x) - \eta(x')| \leq c \, \|x - x'\|_2$

$\Rightarrow$ Geometric term $= c \, \mathbb{E}_{x \sim D, S_{Train}}\left[\| x - \text{nbh}_{S_{Train},1}(x) \|\right]$

$$\boxed{\mathbb{E}_{S_{Train}}\left[\mathcal{L}(f_{S_{Train},1})\right] \leq 2\,\mathcal{L}(f^*) + c\,\mathbb{E}_{x \sim D, S_{Train}}\left[\| x - \text{nbh}_{S_{Tr},1}(x) \|\right]}$$

$$\boxed{\mathbb{E}_{S_{Tr}}\left[\mathcal{L}(f_{S_{Train},1})\right] \leq 2\,\mathcal{L}(f^*) + 4c\sqrt{d}\,(N)^{-\frac{1}{d+1}}}$$

If $N \to \infty$ $\mathbb{E}_{S_{Tr}}\left[\mathcal{L}(f_{S_{Tr},1})\right] \leq 2\,\mathcal{L}(f^*)$

(Proof p12-13, 2066)