

Quiz today 10:45 - 11:15

Please write your full name, ID, email, section (top left)

Announcements

HW2 will be assigned today (due Oct 13th)

Project proposal submission deadline postponed to Nov 3rd

Today:

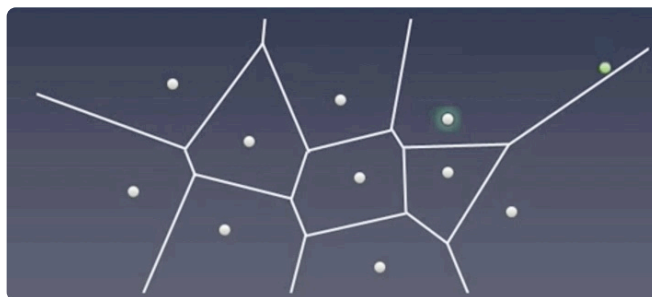
k-Nearest Neighbors

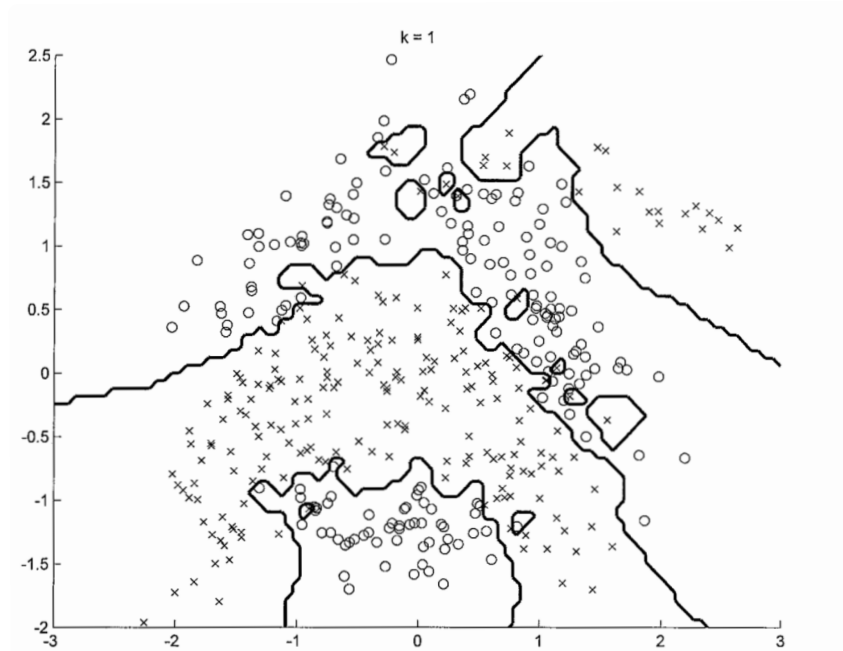
Parametric vs non-parametric models

Nearest Neighbors (NN) cont.

Assign x the same label as the closest training point x_i :

Voronoi

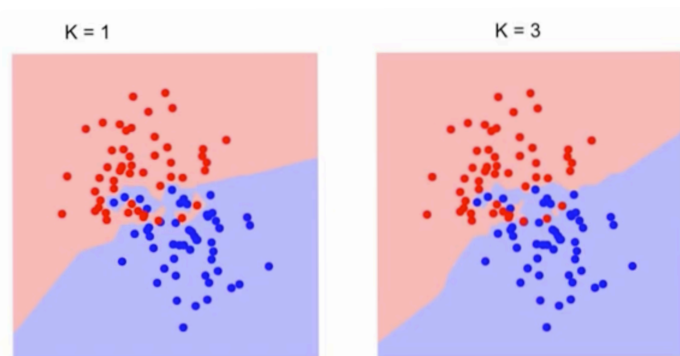




- sensitive to outliers
- no class prior
- no confidence

k-NN

For odd $k \geq 1$, the k -NN rule generalizes the NN rule: Assign x by taking a majority vote over the k training points x_i closest to x



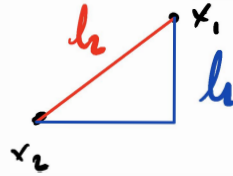
[kNN web demo](#)

$k=1$ is NN rule memorize all examples

$k=n$ always predict the majority class

Which distance functions to use

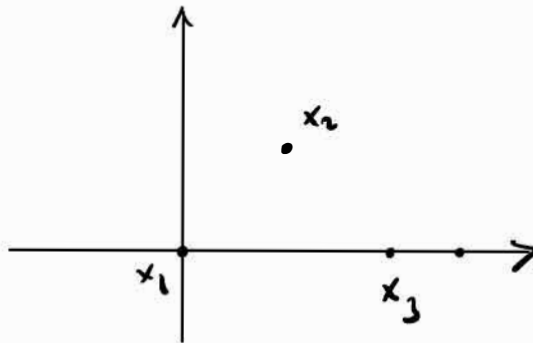
Numerical attributes



l_2 norm

l_1 norm

Other norms



$$\underline{d} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \left\{ \begin{array}{l} 100 \text{ ones} \end{array} \right\}, \quad \underline{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \left\{ \begin{array}{l} 100 \text{ zeros} \end{array} \right\}, \quad \underline{s} = \begin{bmatrix} 100 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(distributed) (spiky)

$$\|\underline{0} - \underline{d}\|_1 = 100$$

$$\|\underline{0} - \underline{d}\|_2 = \sqrt{\sum_{i=1}^{100} 1} = \sqrt{100} = 10$$

$$\|\underline{0} - \underline{s}\|_1 = 100$$

l_2 norm magnifies big entries

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

l_2 norm invariant to rotation $\|Rx\|_2^2 = \langle Rx, Rx \rangle$

l_2 norm depends on the coordinate system $= x^T R^T R x = x^T x = \|x\|_2^2$

Categorical attributes

eg. color : red, green, blue
 0 1 2

Eds. level: high school, bachelor, master, PhD

Hamming distance $d(x, x') = \sum_{i=1}^d 1_{\{x_i \neq x'_i\}}$