

# kNN osztályozás

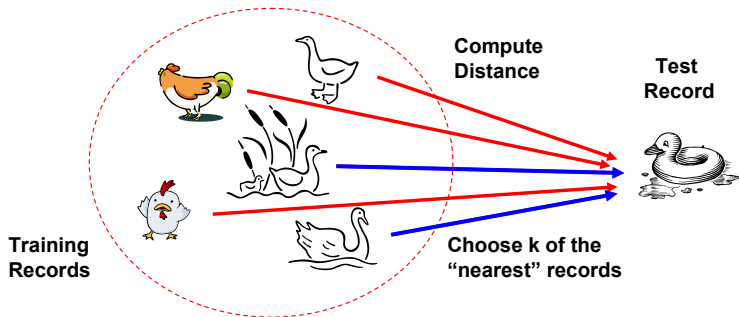
Csima Judit

BME, VIK,  
Számítástudományi és Információelméleti Tanszék

2015. március 25.

# Nearest Neighbor Classifiers

- Basic idea:
  - If it walks like a duck, quacks like a duck, then it's probably a duck

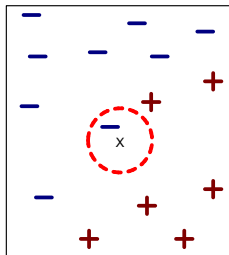


- a rekordok pontok az annyi dimenziós térben, ahány attribútum van (az osztályt nem számítva)
- az osztályozandó sor címkéje a hozzá legközelebb eső  $k$  darab training record alapján lesz valahogyan

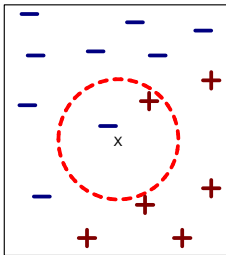
## Mi kell ehhez?

- az összes training record
- milyen távolságot használjunk?
- mi legyen a  $k$ ?

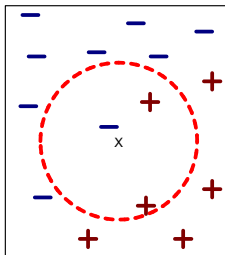
# Definition of Nearest Neighbor



(a) 1-nearest neighbor



(b) 2-nearest neighbor



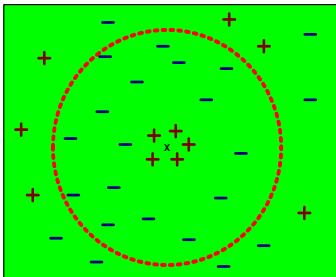
(c) 3-nearest neighbor

K-nearest neighbors of a record  $x$  are data points that have the  $k$  smallest distance to  $x$

- távolság:
  - euklideszi (skálázás kellhet)
  - SMC, Jaccard vagy más, amit tanultunk: az a lényeg, hogy azok legyenek közeliek, akiket annak gondolunk
- döntés a címkéről
  - többségi szavazás a  $k$  szomszéd között
  - súlyozott szavazatok:  $w_i = \frac{1}{d_i^2}$ , ahol a  $d_i$  az  $i$ . szomszéd távolsága
- $k$ 
  - kicsi  $k$  esetén érzékeny a lokális hibákra
  - nagy  $k$  esetén bezavarhat sok távoli

# Nearest Neighbor Classification...

- Choosing the value of  $k$ :
  - If  $k$  is too small, sensitive to noise points
  - If  $k$  is too large, neighborhood may include points from other classes



# Összefoglalás

- lazy learner: csak akkor dolgozik, amikor osztályozandó sor jön
- lassú (drága) egy sor osztályozása, de nincs hosszú előkészítés
- lehet előkészítéssel gyorsan osztályozni, ha  $k = 1$ : felosztjuk a teret cellákra, minden térbeli ponthoz hozzárendeljük a hozzá legközelebbi training recordot: Voronoi-diagramm



# 1 nearest-neighbor

## Voronoi Diagram

