

# Nearest Neighbours

Example 1 An online retailer sells 1000 types of items, and records the profile of a client as

$$v = (x_1, \dots, x_{1000}) \in \mathbb{R}^{1000}$$

$\in x_i =$  amount spent on type  $i$

the retailer is overstocked in three areas

F = furniture

G = garden equipment

H = Holidays

and needs to promote these areas by setting the start page of a given user to precisely one of

page - F

page - G

page - H

for  $v \in \mathbb{R}^{1000}$  define

$f(v) =$  amount spent on area F

$g(v) =$  " " G

$h(v) =$  " " H

Choose training profiles

$P_1, P_2, \dots, P_{600} \in \mathbb{R}^{1000}$  with

large  $f(P_j)$  or  $g(P_j)$  or  $h(P_j)$ .

For a suitable integer parameter,

say  $k=10$ , we determine for

any  $v \in \mathbb{R}^{1000}$  the  $k$  profiles

$P_{j_1}, P_{j_2}, \dots, P_{j_{10}} \in \mathbb{R}^{1000}$  nearest

to  $v$  (for Euclidean metric say).

The start/home page for client  $v$  is set to page- $A$  if the majority of  $P_{j_1}, \dots, P_{j_{10}}$  have large value of  $f(P_j)$ .

This process is called the  $k$ -NN or  $k^{\text{th}}$  nearest neighbour algorithm.

Example 2 A database contains vectors  $p \in \mathbb{R}^{65000}$  representing images of 10 million terrorists/criminals, ten images per terrorist/criminal.

Each traveller at Dublin Airport has their image  $u$  compared to the database. If the Euclidean distance  $\|p - u\|$  is smaller than a fixed threshold  $\epsilon > 0$ , then traveller  $u$  is arrested.

This process is an example of the 1-NN algorithm.

Problem Given a fixed finite set of points  $S \subseteq \mathbb{R}^n$  and an arbitrary  $v \in \mathbb{R}^n$ , how long will it take to find the  $k$  points of  $S$  nearest to  $v$ ?

One solution is to calculate  $\|p-v\|$  for each  $p \in S$  and record those  $p$  that yield the  $k$  smallest values. This takes of the order

$$O(|S|)$$

comparisons, each comparison involving the calculation of

$$\|v-p\| = \sqrt{(x_1-y_1)^2 + \dots + (x_n-y_n)^2}$$

$n$  subtractions,  $n$  squares,  $n$  additions and 1 square root.

At Dublin Airport this is

$$10^8 \times 65000 + 3 \approx 10^{13} \text{ operations,}$$



A better solution involves two steps.

Step 1 Find a linear homomorphism

$$\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

for which  $m$  is small (say  $m \leq 10$ ) and for which there exists a small  $\epsilon > 0$  with

$$(1 - \epsilon)\|v - w\| \leq \|v - \phi(w)\| \leq (1 + \epsilon)\|v - w\|$$

for all  $v, w \in \mathbb{R}^m$ . (e.g. Principal Component Analysis)

Step 2 Solve our problem for small  $n$  (say  $n \leq 10$ ).

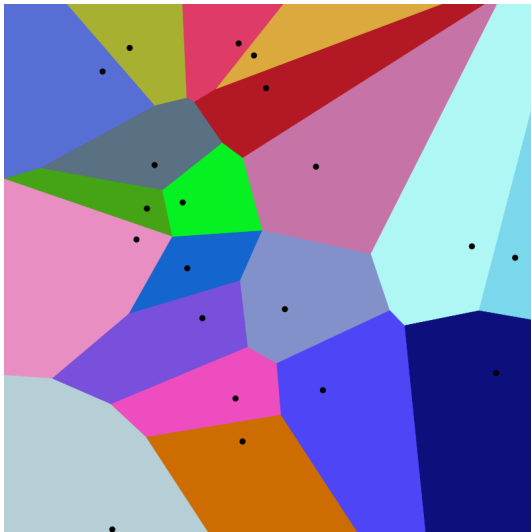
One approach to Step 2 involves Voronoi tessellations of  $\mathbb{R}^m$  based on the finite set  $S \subseteq \mathbb{R}^m$ .

We cut  $\mathbb{R}^n$  into Voronoi  
regions

$$V(p) = \left\{ v \in \mathbb{R}^n : \|p - v\| \leq \|p' - v\| \text{ for} \right. \\ \left. \text{all } p' \in S \right\}$$

one region for each  $p \in S$ .

Voronoi tessellation for set  $S \subset \mathbb{R}^2$  of 20 points.



By Balu Ertl - Own work,  
<https://commons.wikimedia.org/>

Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

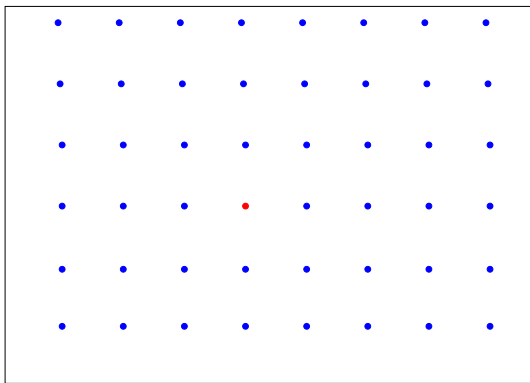


Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

$$V(p) = \{v \in \mathbb{R}^2 : \|p - v\| < \|p' - v\| \text{ for all } v \in S\}$$

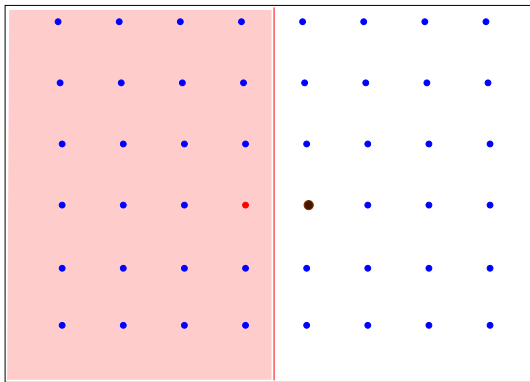
Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

$$V(p) = \{v \in \mathbb{R}^2 : \|p - v\| < \|p' - v\| \text{ for all } v \in S\}$$



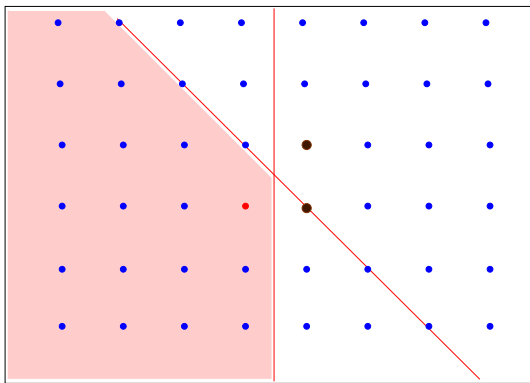
Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

$$V(p) = \{v \in \mathbb{R}^2 : \|p - v\| < \|p' - v\| \text{ for all } v \in S\}$$



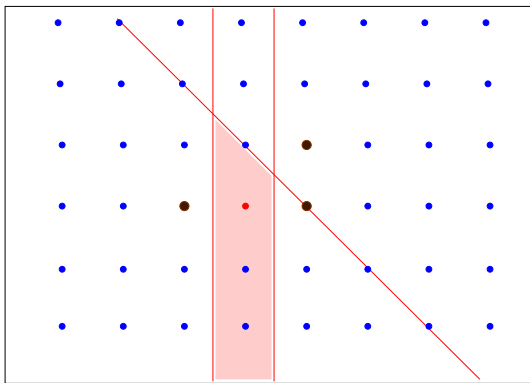
Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

$$V(p) = \{v \in \mathbb{R}^2 : \|p - v\| < \|p' - v\| \text{ for all } v \in S\}$$



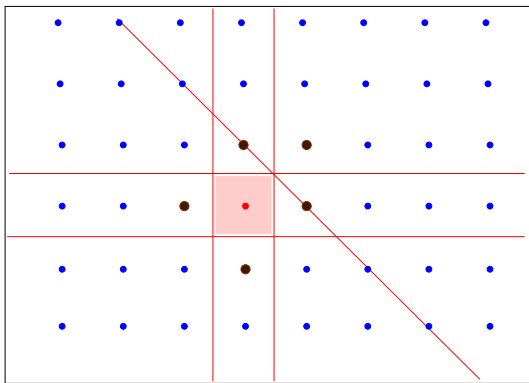
Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

$$V(p) = \{v \in \mathbb{R}^2 : \|p - v\| < \|p' - v\| \text{ for all } v \in S\}$$



Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

$$V(p) = \{v \in \mathbb{R}^2 : \|p - v\| < \|p' - v\| \text{ for all } v \in S\}$$



Consider  $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$ ,  $p = (0, 0) \in S$

$$V(p) = \{v \in \mathbb{R}^2 : \|p - v\| < \|p' - v\| \text{ for all } v \in S\}$$

