Example

$D_{(0)} =$

	1	2	3	4	5
1	0	11	10	14	22
2	11	0	**3**	13	21
3	10	3	0	12	20
4	14	13	12	0	16
5	22	21	20	16	0

$M = 0$

$L = [0]$

merge 2 & 3

$D_{(1)} =$

	1	4	5	6
1	0	14	22	**10**
4	14	0	16	12
5	22	16	0	20
6	10	12	20	0

$M = 1$

$L = [0, 3]$

merge 1 & 6

$D_{(2)} =$

	4	5	7
4	0	16	**12**
5	16	0	20
7	12	20	0

$M = 2$

$L = [0, 3, 10]$

merge 4 & 7

$D_{(3)} =$

	5	8
5	0	16
8	16	0

$m = 3$

$L = [0, 3, 10, 12]$

merge 5 & 8

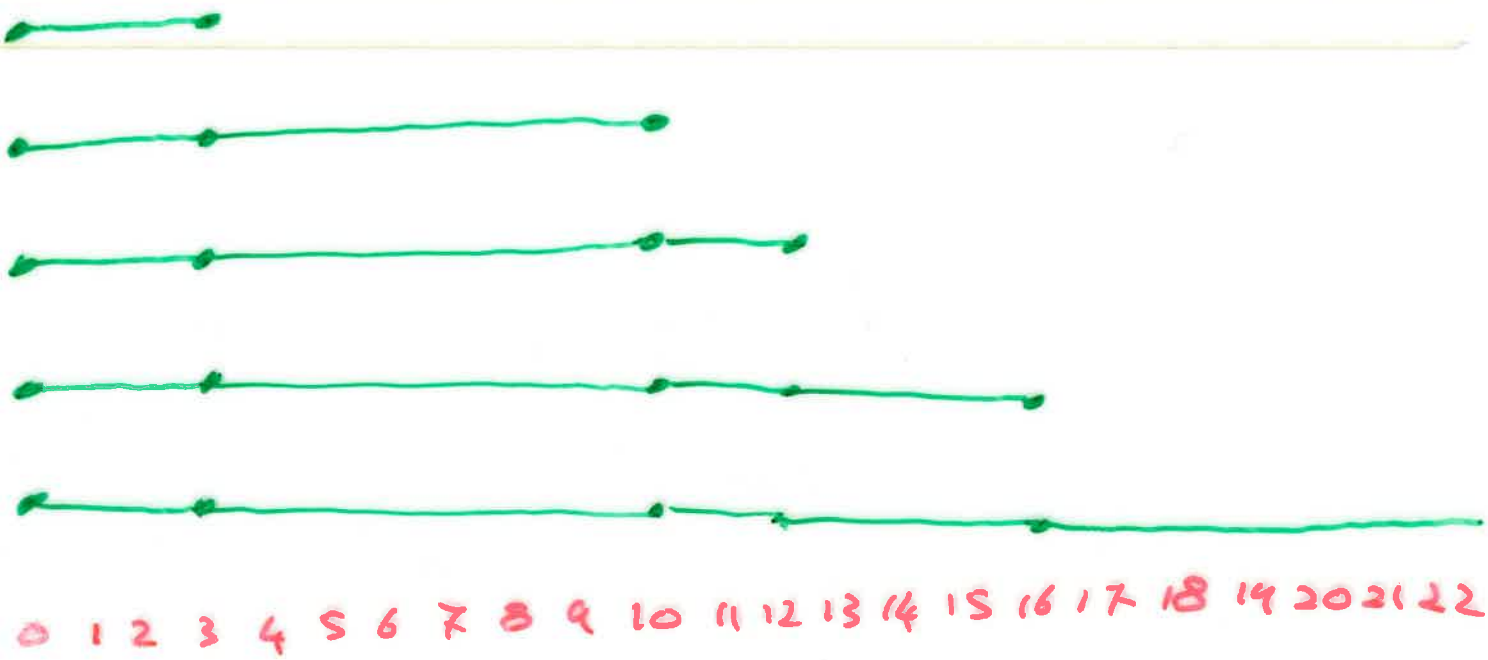
$m = 4$

$D_{(4)} =$

	9
9	

$L = [0, 3, 10, 12, 16]$

Return the bar code



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

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

This process is an example of the 1 nearest neighbour (1-NN) problem.

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?

k-NN problem

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

$$\|v - p\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

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

At Dublin Airport this is

$10^8 \times 65000 \times 3 \simeq 10^{13}$ comparisons.

A better solution involves two steps.

<u>Step 1</u> Find a linear homomorphism

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

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

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

for all $v, w \in \mathbb{R}^n$. (e.g. PCA)

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

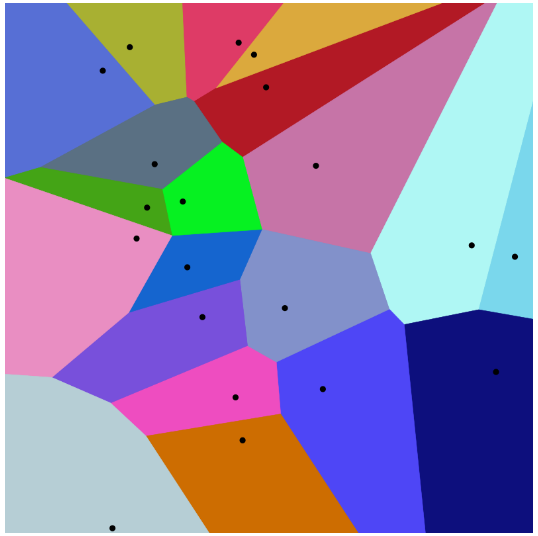
One approach to step 2 involves <u>Voronoi</u> <u>tessellations</u> of $\mathbb{R}^n$ based on the finite database $S \subseteq \mathbb{R}^n$.

We cut $\mathbb{R}^n$ in <u>Voronoi</u> <u>regions</u>

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

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/

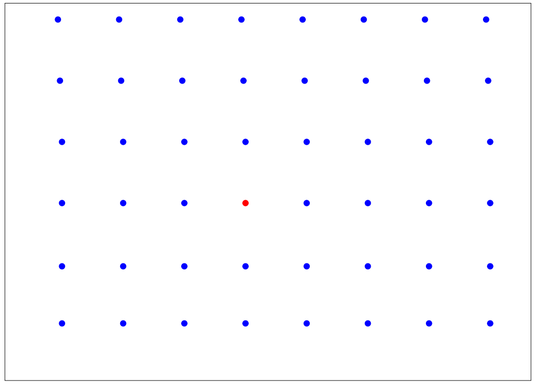
Consider $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2, \quad p = (0,0) \in S$

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

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

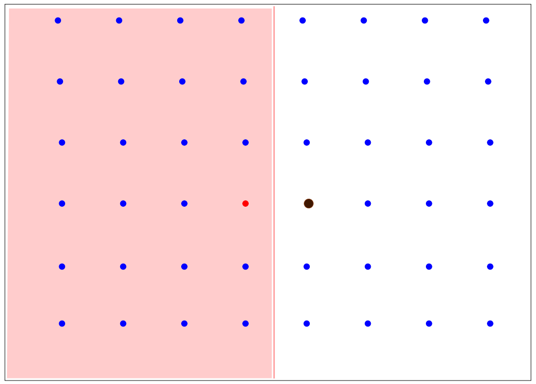
Consider $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2, \quad p = (0,0) \in S$

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



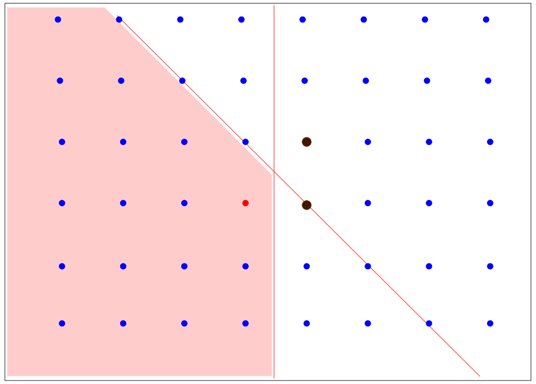
Consider $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2, \quad p = (0,0) \in S$

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



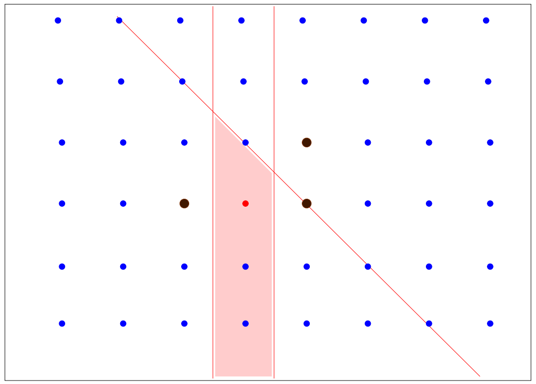
Consider $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2, \quad p = (0,0) \in S$

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



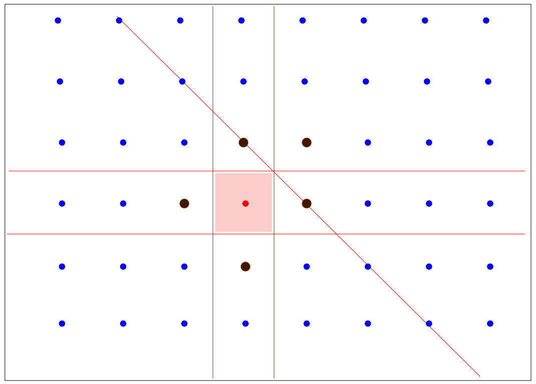
Consider $S = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2, \quad p = (0,0) \in S$

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



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

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



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

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