We have $k=5$, and (x_1, b), (x_2, r), (x_3, b), (x_4, b), (x_5, r) (t_1 → b, t_2 → r, t_3 → b, t_4 → b, t_5 → r)

are the 5 closest neighbors to a test example x.

$$y(x) = \arg\max_{t \in J} \sum_{i=1}^{k} \delta_t(t_i)$$

$t_1 = b$
$t_2 = r$
$t_3 = b$
$t_4 = b$
$t_5 = r$

$J = \{b, r\}$ is the set of labels.

$$y(x) = \arg\max_{t \in \{b,r\}} \sum_{i=1}^{5} 1[t_i = t] = b$$

Case $t=b$ $\Rightarrow (\#\text{votes}) \sum_{i=1}^{5} 1[t_i = b] =$

$= 1[t_1 = b] + 1[t_2 = b] + 1[t_3 = b] + 1[t_4 = b] + 1[t_5 = b]$

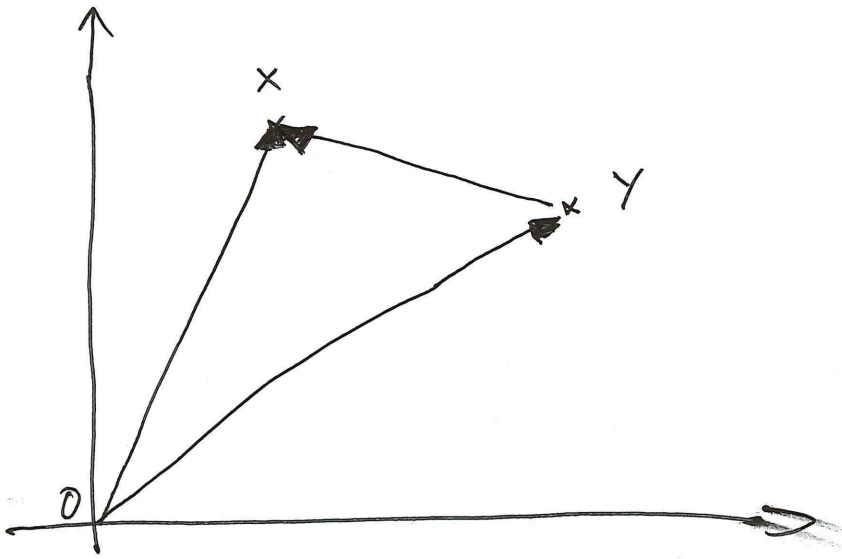
$= 1 + 0 + 1 + 1 + 0$

$= 3$ * max. $\Rightarrow$ arg max $= b$

Case $t=r$ $\Rightarrow (\#\text{votes}) \sum_{i=1}^{5} 1[t_i = r] =$

$= 0 + 1 + 0 + 0 + 1$

$= 2$



$z = x - y$

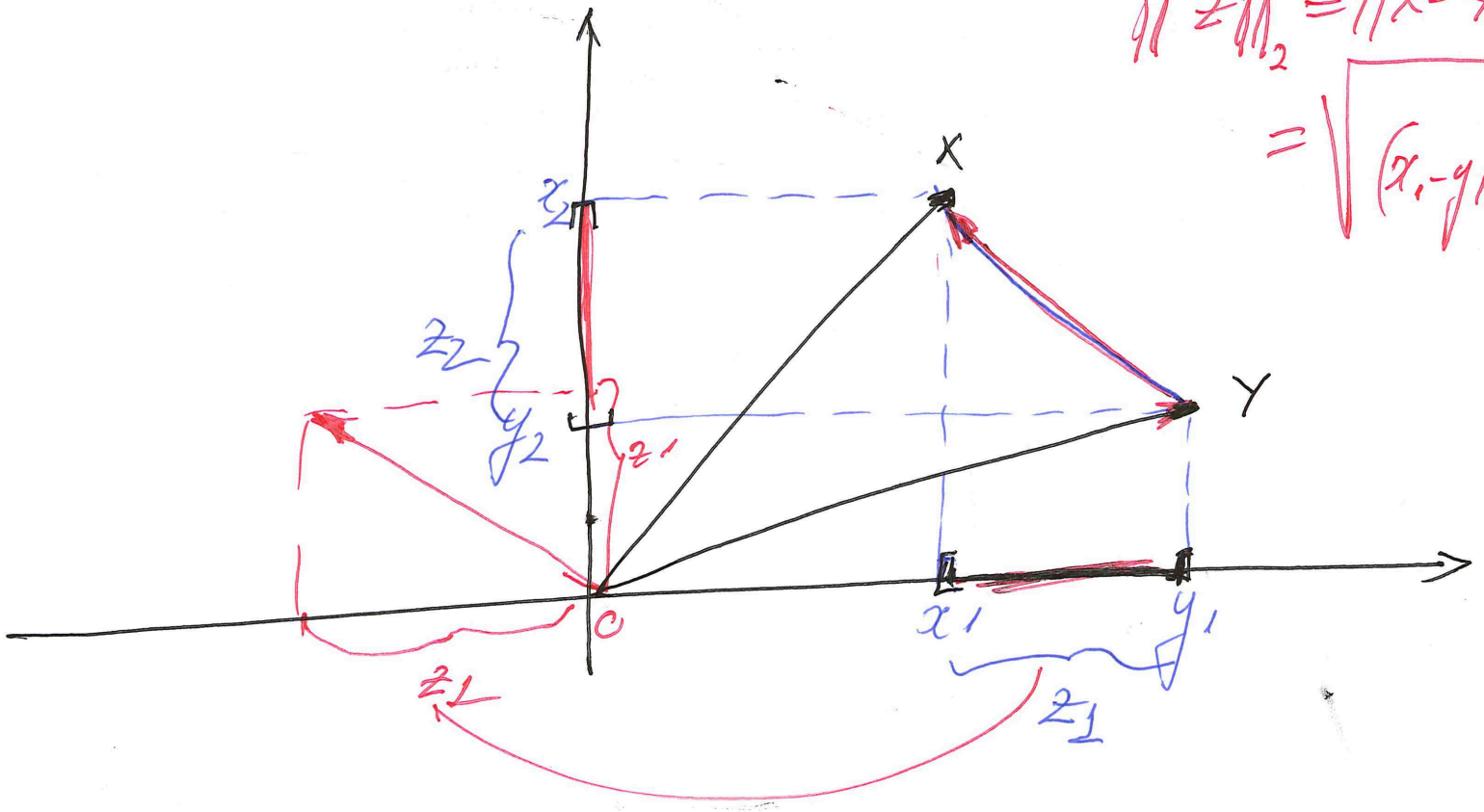
$[x_1, x_2]$ $[y_1, y_2]$

$z = x - y = [x_1 - y_1, x_2 - y_2]$

$= [z_1, z_2]$

$\|z\|_2 = \|x - y\|_2 =$

$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$



Hamming distance between vectors of "discrete" features.

$$x = [0, 1, 1, 0, 1]$$

$$\quad\ \ \neq \ \ = \ \ \neq \ \ = \ \ =$$

$$y = [1, 1, 0, 0, 1]$$

(positions 1, 2, 3, 4, 5)

$$d(x,y) = \sum_{k=1}^{K} 1[x_k \neq y_k] = 2.$$

K → # of features

The edit distance between two strings

the min. number of basic ops (del, ins, or sub) that transforms one string into the other string.

X = trumpf

Y = slump

$d(X, Y) = 3.$

t	r	u	m	p	f
↓ s	↓ s	↓	↓	↓	↓ d
s	l	u	m	p	∅

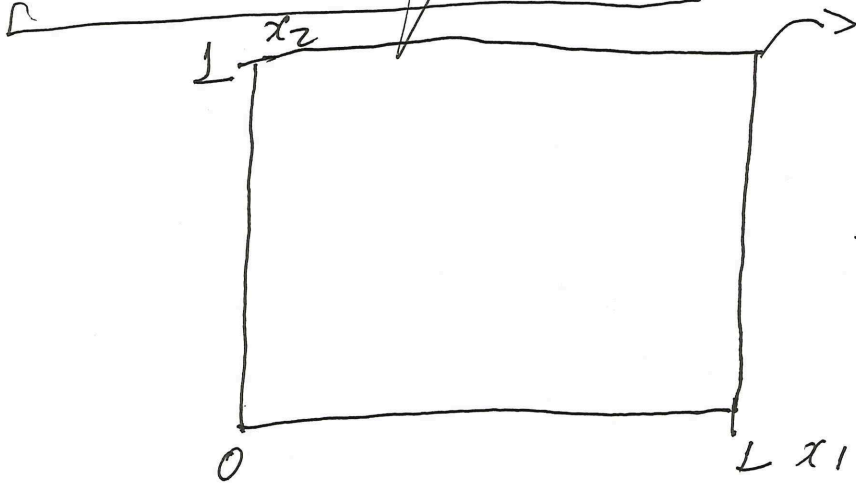
Let's assume that all features are scaled to be in $[0,1]$

Case 1 (one feature): Grid where the "side" of each cell is 0.1.

=> 10 cells => 10 training examples to cover all of them.

Case 2 (two features)

=> unit square



=> 100 cells => need 100 tr. examples to cover.

Case 3 (3 features) => 1,000 cells => 1,000 tr. examples

Case k dim. (k features) => 10^k => exponential # of tr. ex.