Basic notations:

$x = (x_1, x_2, \dots x_n)$. New data points

$z = (z_1, z_2, \dots z_n)$. Training data points

K = Number of nearest neighbours

n: Number of features

* Euclidean Distance

$$d(x,z) = \sqrt{\sum_{i=1}^{n} (x_i - z_i)^2}$$

Example:

Suppose we have two students described by two features

i) study hours

ii) sleep hours

Let, $x = (2,3)$ and $z = (5,7)$

$$d(x,z) = \sqrt{(2-5)^2 + (3-7)^2} = \sqrt{9+16} = \sqrt{25} = 5$$

* Manhattan Distance:

$$d(x,z) = \sum_{i=1}^{n} |x_i - z_i|$$

Example:

$x = (2,3)$, $z = (5,7)$

$$d(x,z) = |2-5| + |3-7| = 3 + 4 = 7$$

* Minkowski Distance

$$d(x,z) = \left(\sum_{i=1}^{n} |x_i - z_i|^p \right)^{1/p}$$

when, $p=1$, it becomes, $d(x,z) = \sum_{i=1}^{n} |x_i - z_i|$ → Mantattan Distance

when, $p=2$, it " , $d(x,z) = \sqrt{\sum_{i=1}^{n} |x_i - z_i|^2}$ → Eucledian Distance

* Cosine Similarity

$$\text{similarity}(x,z) = \frac{x \cdot z}{\|x\| \, \|z\|}$$

Example: $x = (1,0)$, $z_1 = (1,0)$, $z_2 = (0,1)$

Similarity with z_1:

$x \cdot z_1 = 1$ $\|x\| = 1$, $\|z_1\| = 1$

$$\text{similarity}(x, z_1) = \frac{1}{1 \times 1} = 1$$

Similarity with z_2:

$x \cdot z_2 = 0$

$\text{similarity}(x, z_2) = 0$

* Choosing the value of K

Small K: very sensative to noise

Large K: smoother and more stable prediction

Large K: smooth.

$$K \approx \sqrt{N}$$

where N is the number of training samples