K-Nearest Neighbors

Machine Learning Algorithm:

Supervised Learning Model

- Linear Regression
- Logistic Regression/Binary
- Classification
- K-Nearest Neighbors
- Decision Tree

Unsupervised Learning Model

- K-means
- Naive Bayes

Thing that we should choose the K value

1. There is no fixed K value that the K value says will give the best value.

2. We must find those values by ourselves, Maybe compare between 2-3 values at the same time, it is called (Trial and error).

3. A k value that is too low (K=1, K=2) may affect to data set that has an outlier.

4. A k value that is too much maybe over smooth class

Distance Function:

- Calculating between both 2 records for measure similarity of information
- The calculating Distance value must not be negative

Minkowski p=? (p=anything)

Euclidean <----- 2 = p = 1 -----> Manhatton

Euclidean = $De = \sqrt{\sum_{i=2}^{n}(p_i - p_i)^2}$ (p=2)

Manhatton = $D_m = \sum_{i=1}^{n}|p_i - q_i|$ (p=1)

Minkowski = $D = (\sum_{i=1}^{n}|p_i - q_i|^p)^{\frac{1}{p}}$ (p=anything)

Default function: Euclidean

$X = 5, 6, Y = 1, 2$

Euclidean:

$$De = \sqrt{\sum_{i=2}^{n}(p_i - p_i)^2}$$

$$AC^2 = \sqrt{AB^2 + BC^2}$$

$$AC^2 = \sqrt{4^2 + 4^2}$$

$$AC^2 = \sqrt{32}$$

Manhatton:

$$D_m = \sum_{i=1}^{n}|p_i - q_i|$$

$$= (5 - 1) + (6 - 2)$$

$$= 4 + 4$$

$$= 8$$

$$AC^2 = 5.66$$

Adisorn Sriphukthong