

## KNN (K-Nearest Neighbor) classification

步驟:

1) 決定 k 參數

2) 計算距離

歐式距離 (數值資料)  $dE(x_1, x_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

曼哈頓距離 (數值資料)  $dH(x_1, x_2) = |x_1 - x_2| + |y_1 - y_2|$

3) 排序出最小的 k 鄰近物件

4) 根據鄰近物件之結果, 作為判定標準

\* 但非所有資料皆為數值資料, 當計算類別資料

$$\text{距離} = \frac{\text{屬性值相異分數}}{\text{全部屬性分數}}$$

ex: 以 "是否還款" 為例子 若  $k=3$

距離	10	11	12	是否還款
1	$\checkmark \frac{2}{4} = 0.5$	$\frac{4}{4} = 1$	$\checkmark \frac{2}{4} = 0.5$	否
2	$\checkmark 0.5$	$\checkmark 0.5$	1	否
3	0.75	0.75	<u>0.25</u>	是
4	<u>0.25</u>	<u>0.25</u>	0.75	是
5	1	$\checkmark 0.5$	$\checkmark 0.5$	是
6	$\checkmark 0.5$	$\checkmark 0.5$	$\checkmark 0.5$	否
7	<u>0.25</u>	<u>0.25</u>	0.75	是
8	$\checkmark 0.5$	$\checkmark 0.5$	<u>0</u>	否
9	0.75	0.75	0.75	是

\* 距離越近, 影響力可能越大

⇒ 運用 KNN 判斷編號 10, 11, 12 是否還款

$k=3$ , 找出數值前 3 個小的

⑩

0.5 有 2 種結果, 但前 2 個都是 "會還款", 因此估計會還款

⑪

⑫

前 2 個就有 2 種結果, 而 0.5 中較多 "不會還款", 估計不會還款

(0 → 否)  
(0.25 → 是)

	Real	Predict
⑩	否	是
⑪	是	是
⑫	否	否

KNN 分類結果評估：

預測

	是	否
是	1	0
否	1	1

$$\text{Accuracy Rate} = \frac{1+1}{1+0+1+1} = 66.7\%$$

$$\text{Error Rate} = \frac{0+1}{1+0+1+1} = 33.3\%$$

$$\text{Precision} = \frac{1}{1+1} = 50\%$$

$$\text{Recall} = \frac{1}{1+0} = 100\%$$

$$\text{F1-measure} = \frac{2 \times 1}{2 \times 1 + 0 + 1} = 66.7\%$$