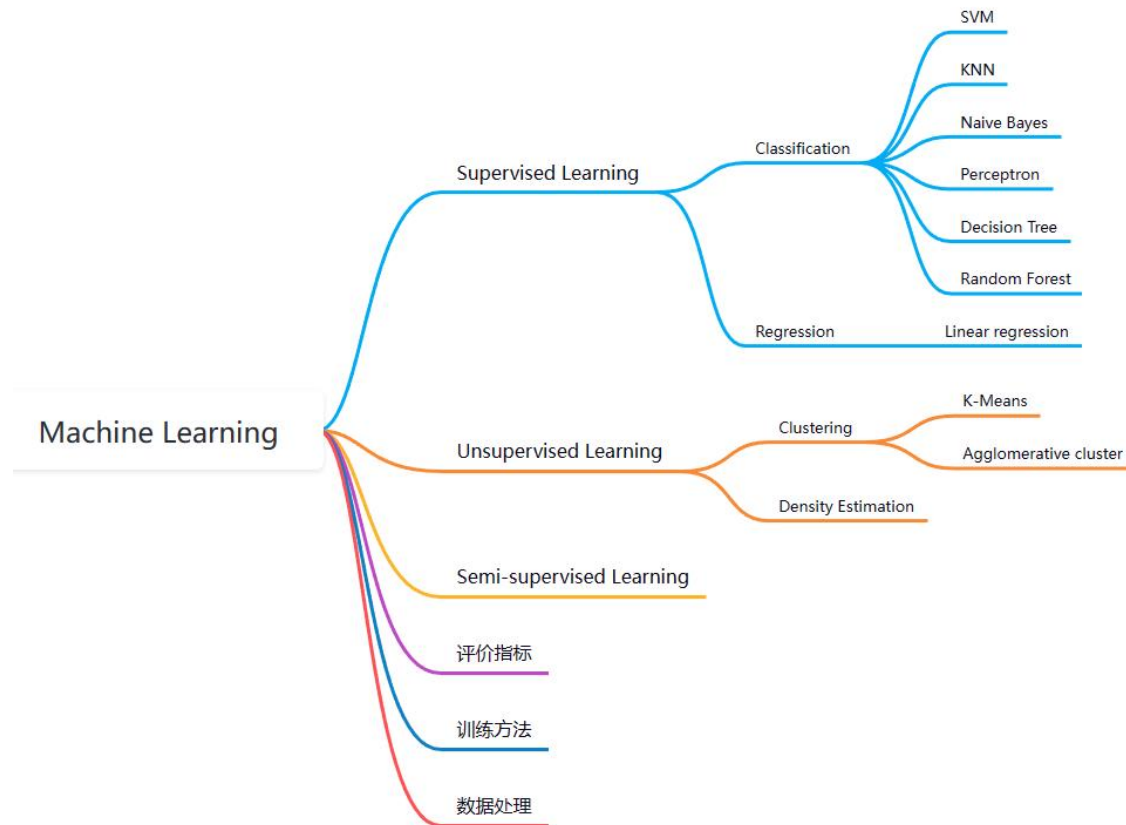


INT104 期末复习



根据往年试题，考试形式不外有监督/无监督模型算法(2021 年)。大题以模型的计算展开。

1.KNN

1. (10 points) The table below shows a training set with 10 examples that is used for training a **3-nearest-neighbors** classifier that uses Manhattan distance, i.e., the distance between two points at coordinates p and q is $|p - q|$. The only attribute, X , is real-valued, and the label Y has two possible classes, 0 and 1. The first fold contains the first 5 examples, and the second fold contains that last 5 examples. In case of ties in distance, use the example with smallest X value as the neighbor. Please compute the 2-fold cross validation accuracy (percentage correct classification).

X	0	1	2	3	4	5	6	7	8	9
Y	1	0	1	0	1	0	1	0	1	0

2.K-means

2. (10 points) You want to cluster 7 points into 3 clusters using the k-means clustering algorithm. Suppose after the first iteration, clusters C_1 , C_2 and C_3 contain the following two-dimensional points:

C_1 contains the 2 points: $\{(0, 6), (6, 0)\}$

C_2 contains the 3 points: $\{(2, 2), (4, 4), (6, 6)\}$

C_3 contains the 2 points: $\{(5, 5), (7, 7)\}$

Please compute the coordinates of **cluster centers** for these 3 clusters.

2. k-means

$$C_1 = \left(\frac{0+6}{2}, \frac{6+0}{2} \right) = (3, 3)$$

$$C_2 = \left(\frac{2+4+6}{3}, \frac{2+4+6}{3} \right) = (4, 4)$$

$$C_3 = \left(\frac{5+7}{2}, \frac{5+7}{2} \right) = (6, 6)$$

3. Naive Bayes (朴素贝叶斯)

3. (20 points) The following dataset as in the table is provided to build a naive Bayes classifier, where $\{x_1, x_2, x_3, x_4\}$ and l are the features and the label, respectively. Please give the process of building the classifier and predict the label of the unknown instance $x = [1, 0, 1, 1]^T$.

x_1	x_2	x_3	x_4	l
0	1	0	1	0
0	0	1	0	0
1	1	1	0	0
1	0	1	0	1

$$3. \quad P(y|x) = \frac{P(y)P(x|y)}{P(x)}$$

$$P(y|x) \propto P(y)P(x|y)$$

\uparrow \uparrow \uparrow
 posterior prior CCP

① when $y = 0$

$$P(y=0) = \frac{3}{4} \quad \text{直接得到}$$

$$P(x_1=0|y=0) = \frac{2}{3}, \quad P(x_1=1|y=0) = \frac{1}{3}$$

$$P(x_2=0|y=0) = \frac{1}{3}, \quad P(x_2=1|y=0) = \frac{2}{3}$$

$$P(x_3=0|y=0) = \frac{1}{3}, \quad P(x_3=1|y=0) = \frac{2}{3}$$

$$P(x_4=0|y=0) = \frac{2}{3}, \quad P(x_4=1|y=0) = \frac{1}{3}$$

因此可以得到:

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

$$P(y=0|x) = \frac{P(y=0) \cdot P(x|y=0)}{P(x)}$$

$$\propto P(y=0) \cdot P(x|y=0)$$

$$= P(x_1=1, x_2=0, x_3=1, x_4=1 | y=0)$$

$$= \prod_{i=1}^4 P(x_i|y=0) = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{81}$$

4.感知机模型

4. (20 points) Perceptron is a function that maps input \mathbf{x} to a label as follows

$$f(\mathbf{x}) = \begin{cases} 1, & w \cdot \mathbf{x} + b > 0 \\ 0, & \text{otherwise} \end{cases}$$

Now consider solving the logical **OR** and logical **XOR** problems (as shown in two tables) with the perceptron model.

$$y = f(\mathbf{x}) = \begin{cases} 1, & w_1x_1 + w_2x_2 + b > 0 \\ 0, & \text{otherwise} \end{cases}$$

Table 1: Logical OR

x_1	x_2	y
0	1	1
1	1	1
1	0	1
0	0	0

Table 2: Logical XOR

x_1	x_2	y
0	1	1
1	1	0
1	0	1
0	0	0

- 1) (4 points) Please draw all datapoints of the tables in the two-dimensional space for logical OR and logical XOR problems, respectively, where different classes are marked with different shapes.
- 2) (16 points) Please explain separately whether the perceptron can mimic the output of logical OR and logical XOR or not. If so, please give an example of function $f(\mathbf{x})$; if not, please prove that there is no such function $f(\mathbf{x})$.

第一题

1. (10 points) The table below shows a training set with 10 examples that is used for training a **3-nearest-neighbors** classifier that uses Manhattan distance, i.e., the distance between two points at coordinates p and q is $|p - q|$. The only attribute, X , is real-valued, and the label Y has two possible classes, 0 and 1. The first fold contains the first 5 examples, and the second fold contains that last 5 examples. In case of ties in distance, use the example with smallest X value as the neighbor. Please compute the 2-fold cross validation accuracy (percentage correct classification).

X	0	1	2	3	4	5	6	7	8	9
Y	1	0	1	0	1	0	1	0	1	0

2020-2021 - final exam

1. KNN: \hat{x} 训练集, x 测试集
 y 为预测值

① $\hat{x} \in [0, 4], x \in [5, 9]$

Input	3-Nearest-neighbors	Predict
$x=5$	$x=2, 3, 4$	$y=1$
$x=6$	$x=2, 3, 4$	$y=1$
$x=7$	$x=2, 3, 4$	$y=1$
$x=8$	$x=2, 3, 4$	$y=1$
$x=9$	$x=2, 3, 4$	$y=1$

The accuracy of this fold is

$$A_1 = \frac{1}{5} \sum_{i=0}^4 I(y_i = \hat{y}_i)$$

$$= \frac{1}{5} (0 + 1 + 0 + 1 + 0)$$

$$= \frac{1}{5} \times 2 = \frac{2}{5} = 40\%$$

② $\hat{x} \in [5, 9], x \in [0, 4]$

Input	3-NN	predict
$x=0$	$x=5, 6, 7$	$y=0$
$x=1$	$x=5, 6, 7$	$y=0$
$x=2$	$x=5, 6, 7$	$y=0$
$x=3$	$x=5, 6, 7$	$y=0$
$x=4$	$x=5, 6, 7$	$y=0$

$$A_2 = \frac{1}{5} \sum_{i=0}^4 I(y_i = \hat{y}_i) = \frac{1}{5} (0 + 1 + 0 + 1 + 0)$$

$$= \frac{2}{5} = 40\%$$

So the total accuracy A is

$$A = \frac{\sum A_i}{n} = \frac{A_1 + A_2}{2} = \frac{\frac{2}{5} + \frac{2}{5}}{2} = \frac{2}{5}$$

$$= 40\%$$

例题

Agglomerative Clustering(层次聚类)

例 14.1 给定 5 个样本的集合, 样本之间的欧氏距离由如下矩阵 D 表示:

$$D = [d_{ij}]_{5 \times 5} = \begin{bmatrix} 0 & 7 & 2 & 9 & 3 \\ 7 & 0 & 5 & 4 & 6 \\ 2 & 5 & 0 & 8 & 1 \\ 9 & 4 & 8 & 0 & 5 \\ 3 & 6 & 1 & 5 & 0 \end{bmatrix}$$

其中 d_{ij} 表示第 i 个样本与第 j 个样本之间的欧氏距离。显然 D 为对称矩阵。应用聚合层次聚类法对这 5 个样本进行聚类。

解 (1) 首先用 5 个样本构建 5 个类, $G_i = \{x_i\}$, $i = 1, 2, \dots, 5$, 这样, 样本之间的距离也就变成类之间的距离, 所以 5 个类之间的距离矩阵亦为 D 。

(2) 由矩阵 D 可以看出, $D_{35} = D_{53} = 1$ 为最小, 所以把 G_3 和 G_5 合并为一个新类, 记作 $G_6 = \{x_3, x_5\}$ 。

(3) 计算 G_6 与 G_1, G_2, G_4 之间的最短距离, 有

$$D_{61} = 2, \quad D_{62} = 5, \quad D_{64} = 5$$

又注意到其余两类之间的距离是

$$D_{12} = 7, \quad D_{14} = 9, \quad D_{24} = 4$$

显然, $D_{61} = 2$ 最小, 所以将 G_1 与 G_6 合并成一个新类, 记作 $G_7 = \{x_1, x_3, x_5\}$ 。

(4) 计算 G_7 与 G_2, G_4 之间的最短距离:

$$D_{72} = 5, \quad D_{74} = 5$$

又注意到

$$D_{24} = 4$$

显然, 其中 $D_{24} = 4$ 最小, 所以将 G_2 与 G_4 合并成一个新类, 记作 $G_8 = \{x_2, x_4\}$ 。

(5) 将 G_7 与 G_8 合并成一个新类, 记作 $G_9 = \{x_1, x_2, x_3, x_4, x_5\}$, 即将全部样本聚成一类, 聚类终止。

上述层次聚类过程可以用图 14.2 所示的层次聚类图表示。

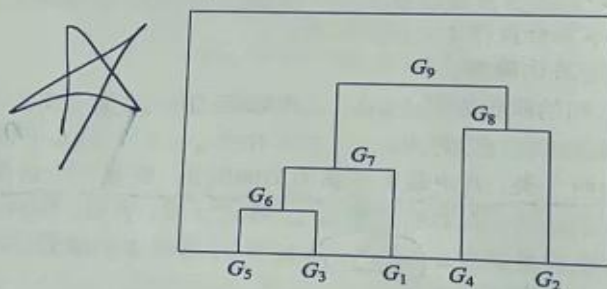


图 14.2 层次聚类图

K-Means

例 14.2 给定含有 5 个样本的集合

$$X = \begin{bmatrix} 0 & 0 & 1 & 5 & 5 \\ 2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

试用 k 均值聚类算法将样本聚到两个类中。

解 按照算法 14.2:

- (1) 选择两个样本点作为类的中心。假设选择 $m_1^{(0)} = x_1 = (0, 2)^T$, $m_2^{(0)} = x_2 = (0, 0)^T$ 。
(2) 以 $m_1^{(0)}, m_2^{(0)}$ 为类 $G_1^{(0)}, G_2^{(0)}$ 的中心, 计算 $x_3 = (1, 0)^T, x_4 = (5, 0)^T, x_5 = (5, 2)^T$ 与 $m_1^{(0)} = (0, 2)^T, m_2^{(0)} = (0, 0)^T$ 的欧氏距离平方。
(a) 对 $x_3 = (1, 0)^T$, $d(x_3, m_1^{(0)}) = 5, d(x_3, m_2^{(0)}) = 1$, 将 x_3 分到类 $G_2^{(0)}$ 。
(b) 对 $x_4 = (5, 0)^T$, $d(x_4, m_1^{(0)}) = 29, d(x_4, m_2^{(0)}) = 25$, 将 x_4 分到类 $G_2^{(0)}$ 。
(c) 对 $x_5 = (5, 2)^T$, $d(x_5, m_1^{(0)}) = 25, d(x_5, m_2^{(0)}) = 29$, 将 x_5 分到类 $G_1^{(0)}$ 。
(3) 得到新的类 $G_1^{(1)} = \{x_1, x_5\}, G_2^{(1)} = \{x_2, x_3, x_4\}$, 计算类的中心 $m_1^{(1)}, m_2^{(1)}$:

$$m_1^{(1)} = (2.5, 2.0)^T, \quad m_2^{(1)} = (2, 0)^T$$

- (4) 重复步骤 (2) 和步骤 (3)。将 x_1 分到类 $G_1^{(1)}$, 将 x_2 分到类 $G_2^{(1)}$, x_3 分到类 $G_2^{(1)}$, x_4 分到类 $G_2^{(1)}$, x_5 分到类 $G_1^{(1)}$, 得到新的类 $G_1^{(2)} = \{x_1, x_5\}, G_2^{(2)} = \{x_2, x_3, x_4\}$ 。

由于得到的新的类没有改变, 聚类停止。得到聚类结果:

$$G_1^* = \{x_1, x_5\}, \quad G_2^* = \{x_2, x_3, x_4\}$$

朴素贝叶斯

例 4.1 试由表 4.1 的训练数据学习一个朴素贝叶斯分类器并确定 $x = (2, S)^T$ 的类标记 y 。表中 $X^{(1)}, X^{(2)}$ 为特征, 取值的集合分别为 $A_1 = \{1, 2, 3\}$, $A_2 = \{S, M, L\}$, Y 为类标记, $Y \in C = \{1, -1\}$ 。

表 4.1 训练数据

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
$X^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1	-1

解 根据算法 4.1, 由表 4.1 容易计算下列概率:

$$P(Y = 1) = \frac{9}{15}, \quad P(Y = -1) = \frac{6}{15}$$

54

$$P(X^{(1)} = 1|Y = 1) = \frac{2}{9}, \quad P(X^{(1)} = 2|Y = 1) = \frac{3}{9}, \quad P(X^{(1)} = 3|Y = 1) = \frac{4}{9}$$

$$P(X^{(2)} = S|Y = 1) = \frac{1}{9}, \quad P(X^{(2)} = M|Y = 1) = \frac{4}{9}, \quad P(X^{(2)} = L|Y = 1) = \frac{4}{9}$$

$$P(X^{(1)} = 1|Y = -1) = \frac{3}{6}, \quad P(X^{(1)} = 2|Y = -1) = \frac{2}{6}, \quad P(X^{(1)} = 3|Y = -1) = \frac{1}{6}$$

$$P(X^{(2)} = S|Y = -1) = \frac{3}{6}, \quad P(X^{(2)} = M|Y = -1) = \frac{2}{6}, \quad P(X^{(2)} = L|Y = -1) = \frac{1}{6}$$

对于给定的 $x = (2, S)^T$, 计算

$$P(Y = 1)P(X^{(1)} = 2|Y = 1)P(X^{(2)} = S|Y = 1) = \frac{9}{15} \cdot \frac{3}{9} \cdot \frac{1}{9} = \frac{1}{45}$$

$$P(Y = -1)P(X^{(1)} = 2|Y = -1)P(X^{(2)} = S|Y = -1) = \frac{6}{15} \cdot \frac{2}{6} \cdot \frac{3}{6} = \frac{1}{15}$$

由于 $P(Y = -1)P(X^{(1)} = 2|Y = -1)P(X^{(2)} = S|Y = -1)$ 最大, 所以 $y = -1$ 。