Artificial Intelligence
——Test and Answer

Yanghui Rao
Assistant Prof., Ph.D
School of Data and Computer Science,
Sun Yat-sen University
raoyangh@mail.sysu.edu.cn

上午 (10:45-12:25)部分

- Given the 5 itemsets, which of the following is false? The minimum support threshold: *minsup* = 40%. (Answer: 4)
  - 1) {A} is a maximal frequent itemset
  - 2) {A} is a closed frequent itemset
  - 3) {A, B} is a closed itemset
  - 4) {B, C} is a closed itemset

ID	Itemsets
I1	A, B
I2	A, C
I3	B, D
I4	A, E, F
I5	A, D

Note: A maximal frequent itemset is defined as a frequent itemset for which none of its immediate supersets are frequent. A closed frequent itemset is defined as a frequent itemset X for which none of its immediate supersets has exactly the same support count as X.

- For the above example, we change it to the following matrix form. If the initial cluster of "I1" is "C1", the initial cluster of "I2" is "C2", the initial cluster of "I3" is "C1", and the initial cluster of "I4" and "I5" is "C2", answer these questions (假设采用基于街区距离的k-Means算法):
- Initially, the centroid of "C1" is ?

  (0.5, 1, 0, 0.5, 0, 0)
- After one iteration, the centroid of "C1" is ?

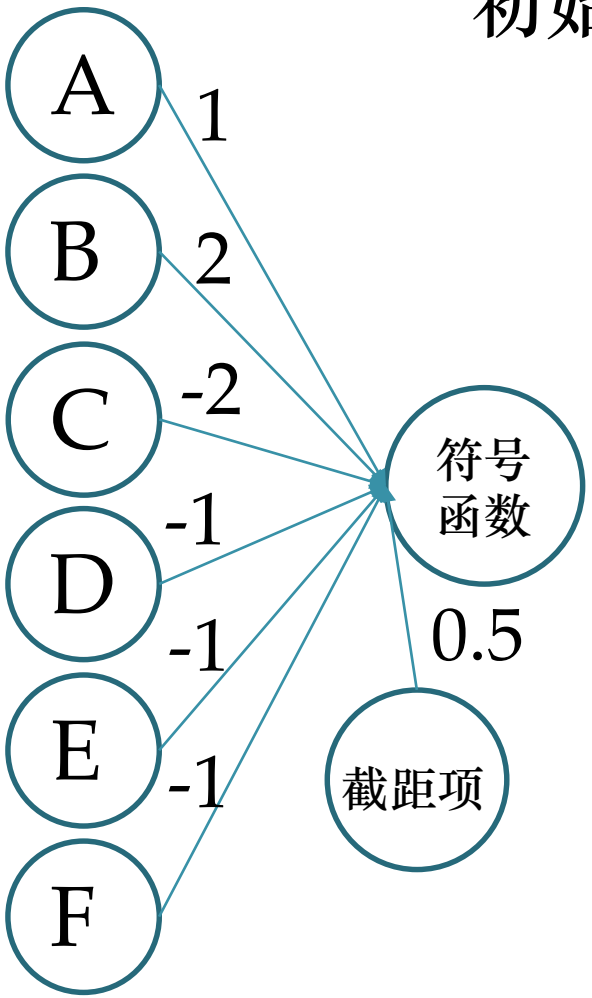
  (0.5, 1, 0, 0.5, 0, 0)

ID	A	B	C	D	E	F
I1	1	1	0	0	0	0
I2	1	0	1	0	0	0
I3	0	1	0	1	0	0
I4	1	0	0	0	1	1
I5	1	0	0	1	0	0

- 给定表格中的训练数据，假设采用PLA算法。初始的各个属性权重如图所示，请计算迭代一次后截距项、A~F的权重分别是多少?

初始权重：(0.5, 1, 2, -2, -1, -1, -1)

答案：(-0.5, 0, 2, -2, -2, -1, -1)



ID	A	B	C	D	E	F	Class Label
I1	1	1	0	0	0	0	+
I2	1	0	1	0	0	0	-
I3	0	1	0	1	0	0	+
I4	1	0	0	0	1	1	-
I5	1	0	0	1	0	0	-

PLA

- O(I1) = sign(0.5+1+2) = 1
- O(I2) = sign(0.5+1-2) = -1
- O(I3) = sign(0.5+2-1) = 1
- O(I4) = sign(0.5+1-1-1) = -1
- O(I5) = sign(0.5+1-1) = 1 -> 错误
- 基于I5，迭代一次后的新权重：

$$\begin{pmatrix} 0.5 \\ 1 \\ 2 \\ -2 \\ -1 \\ -1 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -0.5 \\ 0 \\ 2 \\ -2 \\ -2 \\ -1 \\ -1 \end{pmatrix}$$

- For the above example, predict the class label of feature "A" using the Naïve Bayesian (i.e., NB) classifier. Note: P("B"|Class Label = "+") = 2/4, where 4 is the total number of features within the class label of "+", and 2 is the occurrence number of feature "B" within the class label of "+".
- Answer: "-"

$$P(+|\mathrm{A}) = \frac{P(+)P(\mathrm{A}|+)}{P(\mathrm{A})} = \frac{1}{10P(\mathrm{A})}$$

$$P(-|\mathrm{A}) = \frac{P(-)P(\mathrm{A}|-)}{P(\mathrm{A})} = \frac{9}{35P(\mathrm{A})}$$

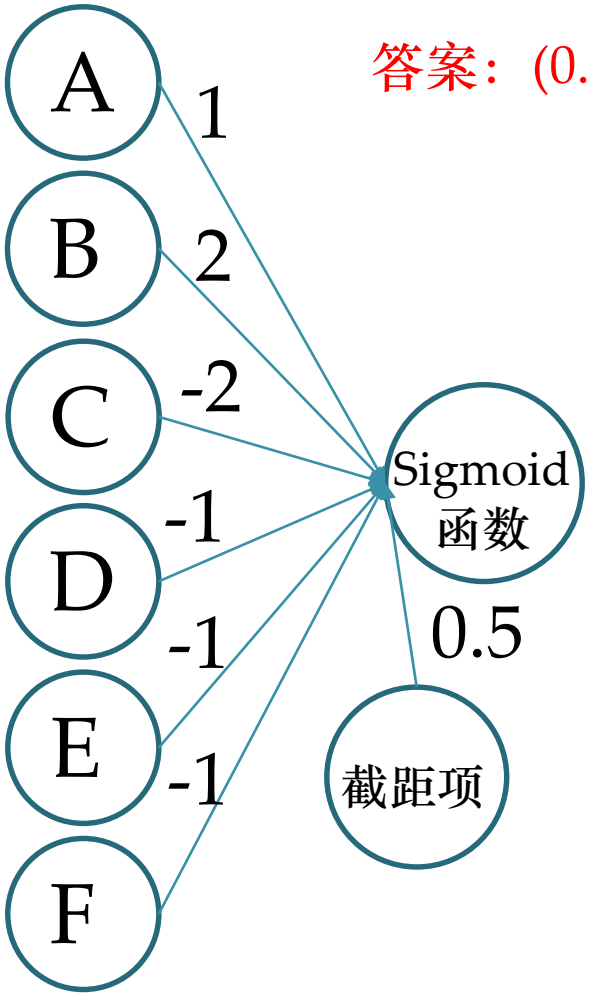
- What's the predicted class label of instance {A, B}? Answer: "+"

$$P(+|\mathrm{A},\mathrm{B}) = \frac{P(+)P(\mathrm{A},\mathrm{B}|+)}{P(\mathrm{A},\mathrm{B})} = \frac{P(+)P(\mathrm{A}|+)P(\mathrm{B}|+)}{P(\mathrm{A},\mathrm{B})}$$

$$P(-|\mathrm{A},\mathrm{B}) = \frac{P(-)P(\mathrm{A},\mathrm{B}|-)}{P(\mathrm{A},\mathrm{B})} = \frac{P(-)P(\mathrm{A}|-)P(\mathrm{B}|-)}{P(\mathrm{A},\mathrm{B})}$$

ID	A	B	C	D	E	F	Class Label
I1	1	1	0	0	0	0	+
I2	1	0	1	0	0	0	-
I3	0	1	0	1	0	0	+
I4	1	0	0	0	1	1	-
I5	1	0	0	1	0	0	-

- 给定表格中的训练数据，假设采用LR算法(学习率=0.1)。初始的各个属性权重如图所示，请计算迭代一次后截距项、A~F的权重分别是多少？初始权重：(0.5, 1, 2, -2, -1, -1, -1)



答案：(0.38, 0.87, 2.02, -2.04, -1.04, -1.04, -1.04)

ID	A	B	C	D	E	F	Class Label
I1	1	1	0	0	0	0	+
I2	1	0	1	0	0	0	-
I3	0	1	0	1	0	0	+
I4	1	0	0	0	1	1	-
I5	1	0	0	1	0	0	-

LR

- $O(I1) = \text{sigmoid}(0.5+1+2) = 1/(1+e^{-3.5})$
- $O(I2) = \text{sigmoid}(0.5+1-2) = 1/(1+e^{0.5})$
- $O(I3) = \text{sigmoid}(0.5+2-1) = 1/(1+e^{-1.5})$
- $O(I4) = \text{sigmoid}(0.5+1-1-1) = 1/(1+e^{0.5})$
- $O(I5) = \text{sigmoid}(0.5+1-1) = 1/(1+e^{-0.5})$
- 迭代一次后的新权重：

$$\begin{pmatrix} 0.5 \\ 1 \\ 2 \\ -2 \\ -1 \\ -1 \\ -1 \end{pmatrix} - 0.1 \begin{pmatrix} (O(I1)-1)+O(I2)+(O(I3)-1)+O(I4)+O(I5) \\ (O(I1)-1)+O(I2)+O(I4)+O(I5) \\ (O(I1)-1)+(O(I3)-1) \\ O(I2) \\ (O(I3)-1)+O(I5) \\ O(I4) \\ O(I4) \end{pmatrix}$$

下午 (14:25-17:55)部分

- 1. Given 5 itemsets, which of the following is false? The minimum support threshold: *minsup* = 40%. (Answer: 3)
  - 1) {A} is a maximal frequent itemset
  - 2) {A} is a closed frequent itemset
  - 3) {B, E} is a closed itemset
  - 4) {A, B} is a closed itemset

ID	Itemsets
I1	A, B
I2	A, C
I3	B, D
I4	A, E, F
I5	A, D

Note: A maximal frequent itemset is defined as a frequent itemset for which none of its immediate supersets are frequent. A closed frequent itemset is defined as a frequent itemset X for which none of its immediate supersets has exactly the same support count as X.

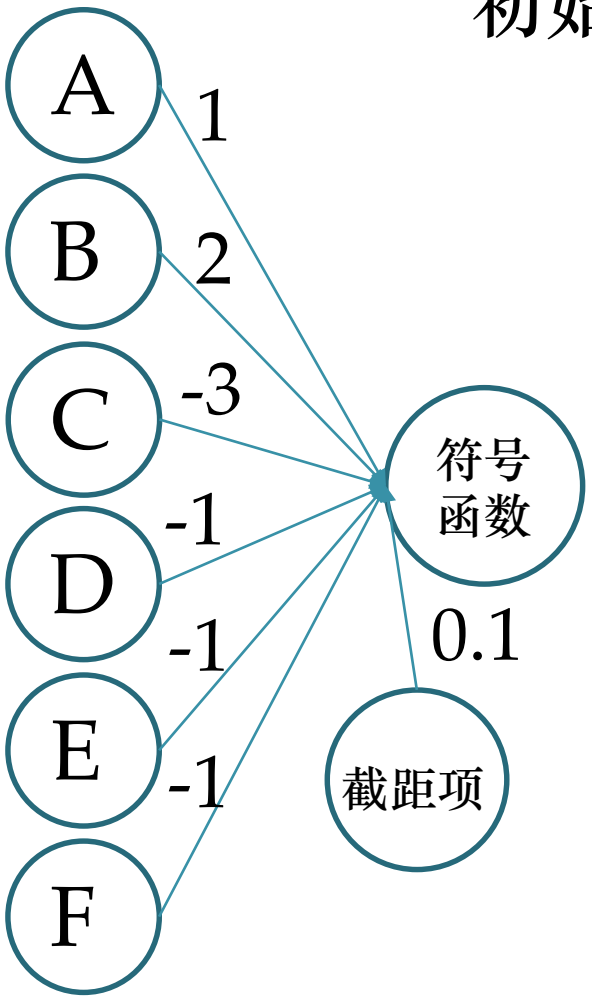
- 2. For the above example, we change it to the following matrix form. If the initial cluster of "I1" is "C1", the initial cluster of "I2" is "C2", the initial cluster of "I3" is "C1", and the initial cluster of "I4" and "I5" is "C2", answer these questions (假设采用基于街区距离的k-Means算法):
- Initially, the centroid of "C2" is ?
  (1, 0, 1/3, 1/3, 1/3, 1/3)
- After one iteration, the centroid of "C2" is ?
  (1, 0, 1/3, 1/3, 1/3, 1/3)

ID	A	B	C	D	E	F
I1	1	1	0	0	0	0
I2	1	0	1	0	0	0
I3	0	1	0	1	0	0
I4	1	0	0	0	1	1
I5	1	0	0	1	0	0

- 3.给定表格中的训练数据，假设采用PLA算法。初始的各个属性权重如图所示，请计算迭代一次后截距项、A~F的权重分别是多少？

初始权重：(0.1, 1, 2, -3, -1, -1, -1)

答案：(-0.9, 0, 2, -3, -2, -1, -1)



ID	A	B	C	D	E	F	Class Label
I1	1	1	0	0	0	0	+
I2	1	0	1	0	0	0	-
I3	0	1	0	1	0	0	+
I4	1	0	0	0	1	1	-
I5	1	0	0	1	0	0	-

PLA

- O(I1) = sign(0.1+1+2) = 1
- O(I2) = sign(0.1+1-3) = -1
- O(I3) = sign(0.1+2-1) = 1
- O(I4) = sign(0.1+1-1-1) = -1
- O(I5) = sign(0.1+1-1) = 1 -> 错误
- 基于I5，迭代一次后的新权重：

$$\begin{pmatrix} 0.1 \\ 1 \\ 2 \\ -3 \\ -1 \\ -1 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -0.9 \\ 0 \\ 2 \\ -3 \\ -2 \\ -1 \\ -1 \end{pmatrix}$$

- 4. For the above example, predict the class label of feature "B" using the Naïve Bayesian (i.e., NB) classifier. Note: P("B"|Class Label = "+") = 2/4, where 4 is the total number of features within the class label of "+", and 2 is the occurrence number of feature "B" within the class label of "+".
- Answer: "+"

$$P(+|\mathrm{B}) = \frac{P(+)P(\mathrm{B}|+)}{P(\mathrm{B})} = \frac{1}{5P(\mathrm{B})}$$

$$P(-|\mathrm{B}) = \frac{P(-)P(\mathrm{B}|-)}{P(\mathrm{B})} = \frac{0}{P(\mathrm{B})}$$

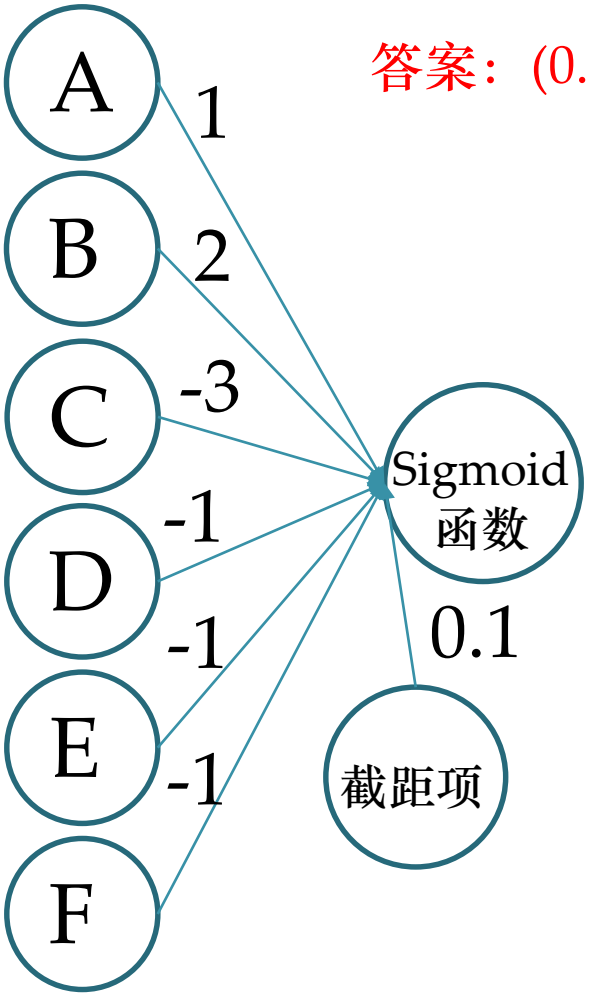
- What's the predicted class label of instance {A, D}? Answer: "-"

$$P(+|\mathrm{A},\mathrm{D}) = \frac{P(+)P(\mathrm{A},\mathrm{D}|+)}{P(\mathrm{A},\mathrm{D})} = \frac{P(+)P(\mathrm{A}|+)P(\mathrm{D}|+)}{P(\mathrm{A},\mathrm{D})}$$

$$P(-|\mathrm{A},\mathrm{D}) = \frac{P(-)P(\mathrm{A},\mathrm{D}|-)}{P(\mathrm{A},\mathrm{D})} = \frac{P(-)P(\mathrm{A}|-)P(\mathrm{D}|-)}{P(\mathrm{A},\mathrm{D})}$$

ID	A	B	C	D	E	F	Class Label
I1	1	1	0	0	0	0	+
I2	1	0	1	0	0	0	-
I3	0	1	0	1	0	0	+
I4	1	0	0	0	1	1	-
I5	1	0	0	1	0	0	-

- 5.给定表格中的训练数据，假设采用LR算法(学习率=0.1)。初始的各个属性权重如图所示，请计算迭代一次后截距项、A~F的权重分别是多少？初始权重：(0.1, 1, 2, -3, -1, -1, -1)



答案：(0.03, 0.91, 2.03, -3.01, -1.03, -1.03, -1.03)

ID	A	B	C	D	E	F	Class Label
I1	1	1	0	0	0	0	+
I2	1	0	1	0	0	0	-
I3	0	1	0	1	0	0	+
I4	1	0	0	0	1	1	-
I5	1	0	0	1	0	0	-

LR

- $O(I1) = \text{sigmoid}(0.1+1+2) = 1/(1+e^{-3.1})$
- $O(I2) = \text{sigmoid}(0.1+1-3) = 1/(1+e^{1.9})$
- $O(I3) = \text{sigmoid}(0.1+2-1) = 1/(1+e^{-1.1})$
- $O(I4) = \text{sigmoid}(0.1+1-1-1) = 1/(1+e^{0.9})$
- $O(I5) = \text{sigmoid}(0.1+1-1) = 1/(1+e^{-0.1})$
- 迭代一次后的新权重：

$$\begin{pmatrix} 0.1 \\ 1 \\ 2 \\ -3 \\ -1 \\ -1 \\ -1 \end{pmatrix} - 0.1 \begin{pmatrix} (O(I1)-1)+O(I2)+(O(I3)-1)+O(I4)+O(I5) \\ (O(I1)-1)+O(I2)+O(I4)+O(I5) \\ (O(I1)-1)+(O(I3)-1) \\ O(I2) \\ (O(I3)-1)+O(I5) \\ O(I4) \\ O(I4) \end{pmatrix}$$