朴素贝叶斯法

- 朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法，是典型的生成学习方法，生成方法由训练数据学习联合概率分布`P(X,Y)`，然后求得后验概率分布`P(Y|X)`。具体来说，利用训练数据学习P(X|Y)和P(Y)的估计，得到联合概率分布：
  $$P(X,Y) = P(Y)P(X|Y)$$
  概率估计方法是极大似然估计或者贝叶斯估计。
- 朴素贝叶斯法的基本假设是条件独立性。
  $$P(X=x,Y=C_k) = P(X^1=x^1,...,X^n=x^n|Y=C_k) = \prod P(X^n=x^n|Y=C_k)$$
- 朴素贝叶斯法利用贝叶斯定理与联合概率模型进行分类预测。
  $$P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{P(Y)P(X|Y)}{\sum P(Y)P(X|Y)}$$

将输入x分到后验概率最大的类y。

$$y = argmaxP(Y=c_k)\prod P(X_j=x^j|Y=c_k)$$

后验概率最大等价于0 -1损失函数时的期望风险最小化。

- 先验概率分布
  $$P(Y=c_k)$$
- 条件概率分布
  $$P(X=x|Y=c_k) = P(X^i=x^i|Y=c_k)$$
- 朴素贝叶斯分类器可表示为
  $$y = argmaxP(Y=c_k)\prod P(X_j=x^j|Y=c_k)$$

后验概率最大化

- 后验概率最大化准则等于期望风险最小化
  $$f(x) = argmaxP(Y=c_k|X=x)$$

朴素贝叶斯的参数估计

- **极大似然估计**
  - 先验概率的极大似然估计
    $$P(Y=c^k) = \frac{\sum I(y_i=c_k)}{N}, k=1,2,...,K$$
  - 条件概率的极大似然估计

$$P(X^j = a_{jl}|Y = c_k) = \frac{\sum I(x_i^j, y_i = c_k)}{\sum I(y_i = c_k)}, j = 1,2,...,n; l = 1,2,...,S; k = 1,2,...,K$$

- **贝叶斯估计**
  - 先验概率的贝叶斯估计，`lambda ≥ 0`，`N`：样本个数，`K`：类别个数。

$$P(Y = c^k) = \frac{\sum I(y_i = c_k) + \lambda}{N + K\lambda}, k = 1,2,...,K$$

  - 条件概率的贝叶斯估计，`Sj`：每个特征值的个数。

$$P(X^j = a_{jl}|Y = c_k) = \frac{\sum I(x_i^j, y_i = c_k) + \lambda}{\sum I(y_i = c_k) + S_j\lambda}$$

- 对于任何`l = 1,2,...,S``k = 1,2,...,K`，有

$$P_\lambda(X^j = a_{jl}|Y = c_k) > 0$$

$$\sum P(X^j = a_{jl}|Y = c_k) = 1$$

例题

- 试由表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X1,X2为特征，取值的集合分别为X1 = {1,2,3}，X2 = {S,M,L}，Y为类标记，Y = {1，-1}。

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
X1	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
X2	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，可以得出

  - 先验概率：

  P(Y=1) = 9/15，P(Y=-1) = 6/15

  - 条件概率：

    P(X1 = 1 | Y = 1) = 2/9，P(X1 = 2 | Y = 1) = 3/9，P(X1 = 3 | Y = 1) = 4/9

    P(X1 = 1 | Y = -1) = 3/6，P(X1 = 2 | Y = -1) = 2/6，P(X1 = 3 | Y = -1) = 1/6

    P(X2 = S | Y = 1) = 1/9，P(X2 = M | Y = 1) = 4/9，P(X2 = L | Y = 1) = 4/9

    P(X2 = S | Y = -1) = 3/6，P(X2 = M | Y = -1) = 2/6，P(X2 = L | Y = -1) = 1/6

- 对于给定的x = (2,S)*T，计算：
  - P(Y = 1)P(X1 = 2 | Y = 1)P(X2 = S | Y = 1) = 1/45
  - P(Y = -1)P(X1 = 2 | Y = -1)P(X2 = S | Y = -1) = 1/15
- 因为`P(Y = -1)P(X1 = 2 | Y = -1)P(X2 = S | Y = -1) = 1/15`最大，所以`y = -1`。