

# 机器学习导论

## 习题五

学号, 作者姓名, 邮箱

2017 年 5 月 31 日

### 1 [25pts] Bayes Optimal Classifier

试证明在二分类问题中, 但两类数据同先验、满足高斯分布且协方差相等时, LDA可产生贝叶斯最优分类器。

**Solution.**

最优贝叶斯分类器的决策面可以由如下公式得到

$$\begin{aligned} P(c_0|\mathbf{x}) &= \frac{P(x|c_0)P(c_0)}{P(c_0)P(x|c_0) + P(c_1)P(x|c_1)} \\ &= \frac{1}{1 + \frac{P(c_1)}{P(c_0)} \exp(-\frac{1}{2}(x - \mu_1)^T \sigma_1^{-1}(x - \mu_1) + \frac{1}{2}(x - \mu_0)^T \sigma_0^{-1}(x - \mu_0))} \\ &= \frac{1}{1 + \exp(x^T \sigma_0^{-1} \mu_1 - x^T \sigma_0^{-1} \mu_0 + \mu_0^T \sigma_0^{-1} \mu_0 - \mu_1^T \sigma_0^{-1} \mu_1)} \\ &= \frac{1}{1 + \exp(b)} \\ P(c_1|\mathbf{x}) &= \frac{P(x|c_1)P(c_1)}{P(c_0)P(x|c_0) + P(c_1)P(x|c_1)} \\ &= \frac{1}{1 + \frac{P(c_0)}{P(c_1)} \exp(\frac{1}{2}(x - \mu_1)^T \sigma_1^{-1}(x - \mu_1) - \frac{1}{2}(x - \mu_0)^T \sigma_0^{-1}(x - \mu_0))} \\ &= \frac{1}{1 + \exp(-x^T \sigma_0^{-1} \mu_1 + x^T \sigma_0^{-1} \mu_0 - \mu_0^T \sigma_0^{-1} \mu_0 + \mu_1^T \sigma_0^{-1} \mu_1)} \\ &= \frac{1}{1 + \exp(-b)} \end{aligned}$$

令 $b=0$ , 得到决策面 $y = \mu_1^T \sigma_0^{-1} x - \mu_0^T \sigma_0^{-1} x$

由LDA可以得到 $w = (S_w^{-1}(\mu_0 - \mu_1) = (2\sigma_0)^{-1}(\mu_0 - \mu_1)$

决策面为 $y = w^T x = \mu_1^T \sigma_0^{-1} x - \mu_0^T \sigma_0^{-1} x$

可以看到两个决策面相同, 命题得证。

### 2 [25pts] Naive Bayes

考虑下面的400个训练数据的数据统计情况, 其中特征维度为2 ( $\mathbf{x} = [x_1, x_2]$ ), 每种特征取值0或1, 类别标记 $y \in \{-1, +1\}$ 。详细信息如表1所示。

根据该数据统计情况，请分别利用直接查表的方式和朴素贝叶斯分类器给出 $\mathbf{x} = [1, 0]$ 的测试样本的类别预测，并写出具体的推导过程。

表 1: 数据统计信息

$x_1$	$x_2$	$y = +1$	$y = -1$
0	0	90	10
0	1	90	10
1	0	51	49
1	1	40	60

**Solution.**

查表得：由于 $x = [1, 0]$ 使 $y$ 为1的样例比 $y$ 为-1的样例多，故预测 $y = 1$

$$P(y = 1) = \frac{271}{400} = 0.6775$$

$$P(y = -1) = \frac{129}{400} = 0.3225$$

$$P(x_1 = 1|y = 1) = \frac{91}{271} = 0.336$$

$$P(x_1 = 1|y = -1) = \frac{109}{129} = 0.845$$

$$P(x_2 = 0|y = 1) = \frac{141}{271} = 0.52$$

$$P(x_2 = 0|y = -1) = \frac{59}{129} = 0.457$$

$$P(y = 1) \times P(x_1 = 1|y = 1) \times P(x_2 = 0|y = 1) = 0.118$$

$$P(y = -1) \times P(x_1 = 1|y = -1) \times P(x_2 = 0|y = -1) = 0.125$$

所以朴素贝叶斯分类器预测为 $y = -1$

### 3 [25pts] Bayesian Network

贝叶斯网(Bayesian Network)是一种经典的概率图模型，请学习书本7.5节内容回答下面的问题：

(1) [5pts] 请画出下面的联合概率分布的分解式对应的贝叶斯网结构：

$$\Pr(A, B, C, D, E, F, G) = \Pr(A) \Pr(B) \Pr(C) \Pr(D|A) \Pr(E|A) \Pr(F|B, D) \Pr(G|D, E)$$

(2) [5pts] 请写出图3中贝叶斯网结构的联合概率分布的分解表达式。

(3) [15pts] 基于第(2)问中的图3，请判断表格2中的论断是否正确，只需将下面的表格填完整即可。

**Solution.** (1) 图见最下方

$$(2) \Pr(A, B, C, D, E, F) = \Pr(A)\Pr(B)\Pr(C|A, B)\Pr(D|B)\Pr(E|C, D)\Pr(F|E)$$

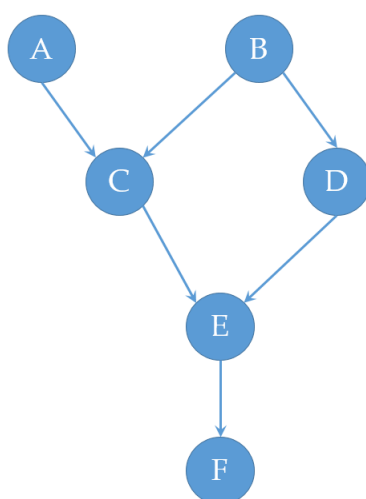


图 1: 题目3-(2)有向图

表 2: 判断表格中的论断是否正确

序号	关系	True/False	序号	关系	True/False
1	$A \perp\!\!\!\perp B$	True	7	$F \perp B C$	False
2	$A \perp B C$	False	8	$F \perp B C, D$	True
3	$C \perp\!\!\!\perp D$	False	9	$F \perp B E$	True
4	$C \perp D E$	False	10	$A \perp\!\!\!\perp F$	False
5	$C \perp D B, F$	False	11	$A \perp F C$	False
6	$F \perp\!\!\!\perp B$	False	12	$A \perp F D$	False

## 4 [25pts] Naive Bayes in Practice

请实现朴素贝叶斯分类器，同时支持离散属性和连续属性。详细编程题指南请参见链接：[http://lamda.nju.edu.cn/ml2017/PS5/ML5\\_programming.html](http://lamda.nju.edu.cn/ml2017/PS5/ML5_programming.html)。

