机器学习导论
习题五

151242041, 王昊庭, hatsuyukiw@gmail.com

2017 年 5 月 30 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. Let the means and the covariance of the two classes be μ_0, μ_1, Σ. The solution of LDA is $\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_0)$. We classify an instance $\mathbf{x}$ as a positive example if

$$\mathbf{w}^{\mathrm{T}} \left(\mathbf{x} - \frac{\mu_0 + \mu_1}{2} \right) \leq 0 \,,$$

and vice versa.

To prove this solution is the optimal Bayesian classifier, we show that the solution outputs $\mathrm{argmax}_c \, P(c|\mathbf{x})$, which is equivalent to $P(c|\mathbf{x}) \geq P(1-c|\mathbf{x})$. Since $P(c|\mathbf{x}) = \frac{P(\mathbf{x}|c)P(c)}{P(\mathbf{x})}$, we only need to prove

$$P(\mathbf{x}|c) \geq P(\mathbf{x}|1-c) \,.$$

We expand this into

$$\frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mu_c)^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \mu_c)\right)}{\sqrt{|2\pi\Sigma|}} \geq \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mu_{1-c})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \mu_{1-c})\right)}{\sqrt{|2\pi\Sigma|}} \,.$$

After some algebraic manipulation we arrive at

$$(\mu_1 - \mu_0)^{\mathrm{T}} \Sigma^{-1} (2\mathbf{x} - \mu_c - \mu_{1-c}) \leq 0 \,.$$

This is exactly

$$\mathbf{w}^{\mathrm{T}} \left(\mathbf{x} - \frac{\mu_0 + \mu_1}{2} \right) \leq 0 \,,$$

when $c = 1$, and vice versa.

Thus, LDA gives the optimal Bayesian classifier.

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. (1) Table lookup:

$$P(y = +1|\mathbf{x} = [0, 0]) = 0.9\,,$$
$$P(y = +1|\mathbf{x} = [0, 1]) = 0.9\,,$$
$$P(y = +1|\mathbf{x} = [1, 0]) = 0.51\,,$$
$$P(y = +1|\mathbf{x} = [1, 1]) = 0.4\,.$$

the classifier either outputs the above probabilities or the following decisions,

$$c(\mathbf{x} = [0, 0]) = +1\,,$$
$$c(\mathbf{x} = [0, 1]) = +1\,,$$
$$c(\mathbf{x} = [1, 0]) = +1\,,$$
$$c(\mathbf{x} = [1, 1]) = -1\,.$$

(2) Naive Bayesian Classifier: First, work out the priors,

$$P(y = +1) = \frac{271}{400}\,,$$
$$P(y = -1) = \frac{129}{400}\,.$$

Then, the posteriors,

$$
\begin{aligned}
P(x_1 = 1|y = +1) &= \frac{91}{271}\,, \\
P(x_1 = 0|y = +1) &= \frac{180}{271}\,, \\
P(x_2 = 1|y = +1) &= \frac{130}{271}\,, \\
P(x_2 = 0|y = +1) &= \frac{141}{271}\,, \\
P(x_1 = 1|y = -1) &= \frac{109}{129}\,, \\
P(x_1 = 0|y = -1) &= \frac{20}{129}\,, \\
P(x_2 = 1|y = -1) &= \frac{70}{129}\,, \\
P(x_2 = 0|y = -1) &= \frac{59}{129}\,.
\end{aligned}
$$

$$
\begin{aligned}
P(y = +1)P(x_1 = 1|y = +1)P(x_2 = 1|y = +1) &= \frac{11830}{108400}\,, \\
P(y = -1)P(x_1 = 1|y = -1)P(x_2 = 1|y = -1) &= \frac{7630}{51600}\,, \\
P(y = +1)P(x_1 = 1|y = +1)P(x_2 = 0|y = +1) &= \frac{12831}{108400}\,, \\
P(y = -1)P(x_1 = 1|y = -1)P(x_2 = 0|y = -1) &= \frac{6431}{51600}\,, \\
P(y = +1)P(x_1 = 0|y = +1)P(x_2 = 1|y = +1) &= \frac{23400}{108400}\,, \\
P(y = -1)P(x_1 = 0|y = -1)P(x_2 = 1|y = -1) &= \frac{1400}{51600}\,, \\
P(y = +1)P(x_1 = 0|y = +1)P(x_2 = 0|y = +1) &= \frac{25380}{108400}\,, \\
P(y = -1)P(x_1 = 0|y = -1)P(x_2 = 0|y = -1) &= \frac{1180}{51600}\,.
\end{aligned}
$$

The classifier outputs

$$
\begin{aligned}
&\operatorname*{argmax}_c P(y = c|\mathbf{x} = [0,0]) = +1\,, \\
&\operatorname*{argmax}_c P(y = c|\mathbf{x} = [0,1]) = +1\,, \\
&\operatorname*{argmax}_c P(y = c|\mathbf{x} = [1,0]) = -1\,, \\
&\operatorname*{argmax}_c P(y = c|\mathbf{x} = [1,1]) = -1\,.
\end{aligned}
$$

To summarize, the look-up method gives $c(\mathbf{x} = [1,0]) = +1$, and the Bayesian classifier gives $c(\mathbf{x} = [1,0]) = -1$. This divergence comes from the Bayesian assumption that the attributes are independent.

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。

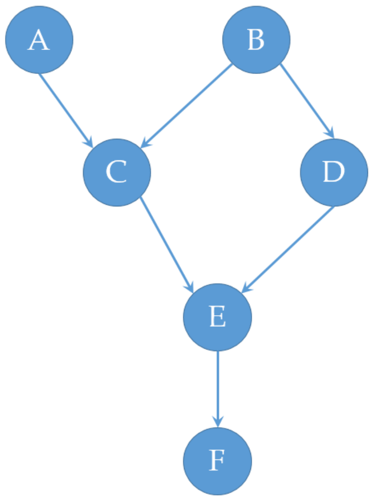

图 1: 题目 3-(2) 有向图

(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。

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	$A \perp\!\!\!\perp B$	T	7	$F \perp\!\!\!\perp B \| C$	F
2	$A \perp\!\!\!\perp B \| C$	F	8	$F \perp\!\!\!\perp B \| C, D$	T
3	$C \perp\!\!\!\perp D$	F	9	$F \perp\!\!\!\perp B \| E$	T
4	$C \perp\!\!\!\perp D \| E$	F	10	$A \perp\!\!\!\perp F$	F
5	$C \perp\!\!\!\perp D \| B, F$	F	11	$A \perp\!\!\!\perp F \| C$	F
6	$F \perp\!\!\!\perp B$	F	12	$A \perp\!\!\!\perp F \| D$	F

Solution. (1)

(2)

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

(3) The method to identify conditional and marginal independent pairs is in http://web.mit.edu/jmn/www/6.034/d-separation.pdf.

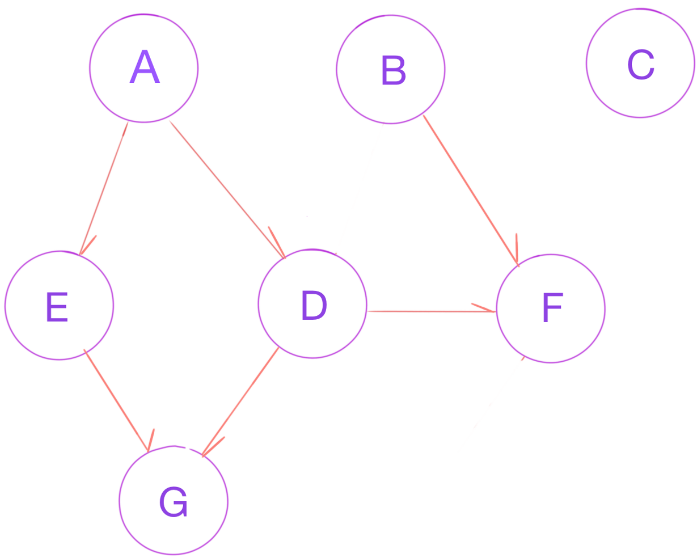

图 2: 3-(1) 答案

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.

同时，请简要谈谈你的感想。实践过程中遇到了什么问题，你是如何解决的？

Solution. 朴素贝叶斯分类器的性能在属性间的相关性较大时性能不佳.

需要注意的地方是 Laplacian 修正, 以及样本标准差的计算公式分母是 $n-1$.

遇到的问题有, 在数据集中, 连续属性存在大量的标准差为 0 的情况, 需要特殊处理. 在这种情况下, 正态分布退化为单点分布. 这使得属性值不等于均值的情况下所得概率为 0. 由于这种情况特别频繁, 导致对于事实上测试集中很多数据, 朴素贝叶斯分类器认为数据属于 5 个类别的概率均为 0, 此时分类器只能随机输出一个结果. 这是分类器在给定的数据集上性能不佳的一个原因.