机器学习导论
习题五

学号：141130077
作者姓名：邱梓豪
邮箱：2957606241@qq.com

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. 此处用于写证明(中英文均可)
先求这种条件下的LDA分类器。因为两类数据满足高斯分布，所以有：

$$x|y = k \sim N(\mu_k, \Sigma) \quad k = 0, 1 \tag{1.1}$$

其中：$N(\mu_k, \Sigma)$为多变量高斯分布，均值为μ_k，协方差为μ_k；x$\in \mathbb{R}^d$
故x的概率函数为：

$$\phi(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}\Sigma^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \tag{1.2}$$

在LDA中，设每一类的先验概率为π_k，均值向量为μ_k，协方差矩阵为Σ，假设共有n个样本，于是可根据训练集对参数进行如下估计：

$$\hat{\pi_k} = \frac{|\{i|y_i = k\}|}{n} \quad k = 0, 1 \tag{1.3}$$

$$\hat{\mu_k} = \frac{1}{|\{i|y_i = k\}|} \sum_{i:y_i=k} x_i \quad k = 0, 1 \tag{1.4}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=0}^{1} \sum_{i:y_i=k} (x_i - \hat{\mu_k})(x_i - \hat{\mu_k})^T \tag{1.5}$$

设关于x的估计判别函数（estimated discriminant function）为：

$$\delta_k(x) = \hat{\pi_k}\phi(x; \hat{\mu_k}, \hat{\Sigma}) \tag{1.6}$$

所以，LDA分类器变为：

$$\begin{aligned}
\hat{f}^{LDA}(x) &= \arg\max_k \delta_k(x) &(1.7)\\
&= \arg\max_k \hat{\pi_k}\phi(x;\hat{\mu_k},\hat{\Sigma}) &(1.8)\\
&= \arg\max_k \log\hat{\pi_k} + \log\phi(x;\hat{\mu_k},\hat{\Sigma}) &(1.9)\\
&= \arg\max_k -\frac{1}{2}(x-\hat{\mu_k})^T\Sigma^{-1}(x-\hat{\mu_k}) &(1.10)
\end{aligned}$$

上面能得出最后一个式子是因为两类同先验且协方差相等。
再看贝叶斯最优分类器，其表达式如下：

$$h^*(x) = \arg\max_k P(y_k|x) \quad k=0,1 \tag{1.11}$$

由贝叶斯定理可知：

$$P(y_k|x) = \frac{P(x|y_k)P(y_k)}{P(x)} \propto P(x|y_k)P(y_k) \tag{1.12}$$

所以有：

$$\begin{aligned}
h^*(x) &= \arg\max_k P(x|y_k)P(y_k) &(1.13)\\
&= \arg\max_k \log P(x|y_k) + \log P(y_k) &(1.14)\\
&= \arg\max_k \log P(x|y_k) &(1.15)\\
&= \arg\max_k \log\frac{1}{(2\pi)^{\frac{d}{2}}\Sigma^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] &(1.16)\\
&= \arg\max_k -\frac{1}{2}(x-\hat{\mu_k})^T\Sigma^{-1}(x-\hat{\mu_k}) &(1.17)
\end{aligned}$$

上面（1.14）到（1.15）是因为两类同先验；（1.16）到（1.17）是因为两类协方差相等。
故有LDA导出的结果（1.10）和贝叶斯最优分类器导出的结果（1.17）相同，因此LDA可以产生贝叶斯最优分类器。

注：我这里推导的LDA和书上的Fisher判别分析略有差别，但可以从这里的LDA推出书上的Fisher判别分析，推导如下：
注意到式（1.10）：

$$\hat{f}^{LDA}(x) = \arg\max_k -\frac{1}{2}(x-\hat{\mu_k})^T\Sigma^{-1}(x-\hat{\mu_k})$$

对Σ进行特征值分解可得：

$$\hat{\Sigma} = UDU^T \tag{1.18}$$

其中U为标准正交矩阵，$D = diag(a_1, a_2, \cdots, a_d)$，则有：

$$\hat{\Sigma}^{-1} = UD^{-1}U^T \tag{1.19}$$

所以有：

$$
\begin{aligned}
(x-\hat{\mu_k})\Sigma^{-1}(x-\hat{\mu_k})^T &= (x-\hat{\mu_k})^T U D^{-1} U^T (x-\hat{\mu_k}) && (1.20)\\
&= [U^T(x-\hat{\mu_k})]^T D^{-1} [U^T(x-\hat{\mu_k})] && (1.21)\\
&= [D^{\frac{1}{2}} U^T(x-\hat{\mu_k})]^T [D^{\frac{1}{2}} U^T(x-\hat{\mu_k})] && (1.22)\\
&= |D^{\frac{1}{2}} U^T x - D^{\frac{1}{2}} U^T \hat{\mu_k}|^2 && (1.23)\\
&= |\tilde{x} - \tilde{\hat{\mu_k}}|^2 && (1.24)
\end{aligned}
$$

所以有：

$$\hat{f}^{LDA}(x) = |\tilde{x} - \tilde{\hat{\mu_k}}|^2 \tag{1.25}$$

若去掉这里的两类样本先验概率相同，协方差相同的限制，则上式就变成了Fisher判别分析。

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. 此处用于写解答(中英文均可)

（1）直接查表：

$P(y=1|x_1=1,x_2=0)=\frac{51}{51+49}=0.51$

$P(y=-1|x_1=1,x_2=0)=\frac{49}{51+49}=0.49$

所以根据直接查表的结果可预测样本的类别为1。

（2）朴素贝叶斯分类器：

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

P$(x_1=1|y=1)$=$\frac{91}{271}$ P$(x_1=1|y=-1)$=$\frac{109}{129}$

P$(x_2=0|y=1)$=$\frac{141}{271}$ P$(x_2=0|y=-1)$=$\frac{59}{129}$

所以有：样本类别为1的概率=P(y=1)P$(x_1=1|y=1)$P$(x_2=0|y=1)$=0.118

样本类别为-1的概率=P(y=-1)P$(x_1=1|y=-1)$P$(x_2=0|y=-1)$=0.124

所以根据朴素贝叶斯分类的结果可预测样本的类别为-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：

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

(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。

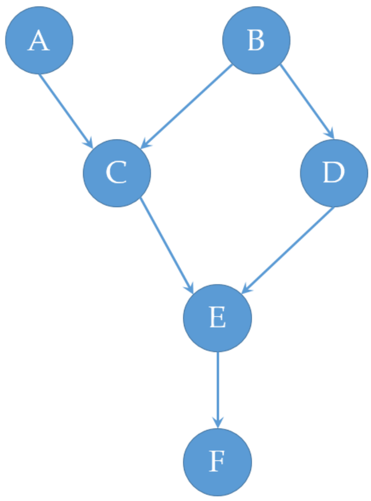

图 1: 题目3-(2)有向图

(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	$A \perp\!\!\!\perp B$	true	7	$F \perp B \mid C$	false
2	$A \perp B \mid C$	false	8	$F \perp B \mid C, D$	true
3	$C \perp\!\!\!\perp D$	false	9	$F \perp B \mid E$	true
4	$C \perp D \mid E$	false	10	$A \perp\!\!\!\perp F$	false
5	$C \perp D \mid B, F$	false	11	$A \perp F \mid C$	false
6	$F \perp\!\!\!\perp B$	false	12	$A \perp F \mid D$	false

Solution. 此处用于写解答(中英文均可)

（1）

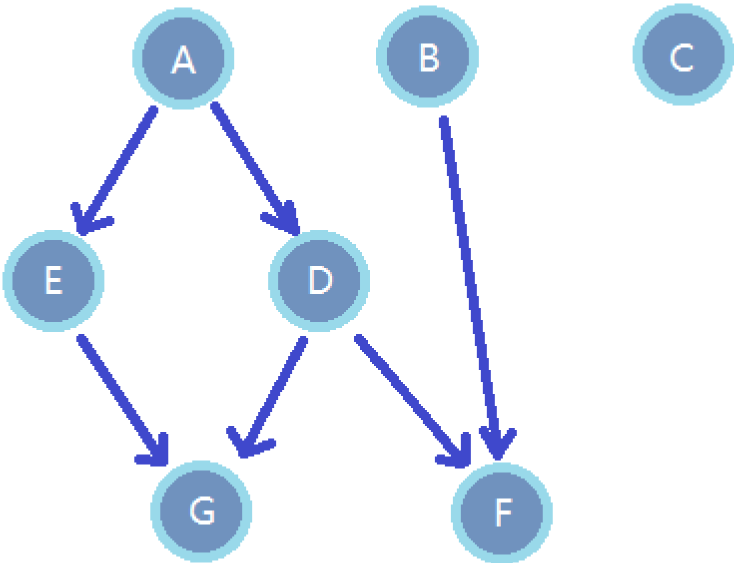

图 2: 题目3-(1)贝叶斯网

（2）分解表达式如下：

$P(A,B,C,D,E,F) = P(A)P(B)P(C|A,B)P(D|B)P(E|C,D)P(F|E)$

（3）

1.$A \perp\!\!\!\perp B$正确。当C不确定时，A和B边缘独立。

2.$A \perp B|C$错误。当C确定时，A和B不相互独立。

3.$C \perp\!\!\!\perp D$错误。当B未知时，C和D不相互独立。

4.$C \perp D|E$错误。理由同上。

5.$C \perp D|B,F$错误。由道德图可知BF已知时CD不相互独立。

6.$F \perp\!\!\!\perp B$错误，顺序结构，中间变量未知，B和F不相互独立。

7.$F \perp B|C$错误，通过原图的道德图可以看出，B和F在C已知的情况下不相互独立。

8.$F \perp B|C,D$正确，通过原图的道德图可以看出，B和F在CD已知的情况下相互独立。

9.$F \perp B|E$正确，通过原图的道德图可以看出，B和F在E已知的情况下相互独立。

10.$A \perp\!\!\!\perp F$错误，顺序结构，中间变量未知，A和F不相互独立。

11.$A \perp F|C$错误，通过原图的道德图可以看出，A和F在C已知的情况下不相互独立。

12.$A \perp F|D$错误，通过原图的道德图可以看出，A和F在D已知的情况下不相互独立。

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. 此处用于写解答(中英文均可)
在实践过程中我遇到的主要问题就是对连续属性的处理。我在处理时发现不少连续属性的方差为0，这这让我很犹豫，最后的的做法是忽略这些属性，因为根据我的观察，我发想这些属性方差为0的原因在于这些属性上的取值中非0值个数太少，我觉得这时这个属性已经不适合作为一个分类的依据，所以我忽略了这些方差为0的属性。