深度之眼
deepshare.net

# 法律声明

deepshare.net
深度之眼

# 西瓜书公式推导

导师：Sm1les（Datawhale南瓜书项目负责人）

# 朴素贝叶斯公式推导

Derivation of Naive Bayes formula

# 本节大纲

Outline



先修内容：西瓜书7.1、7.2、7.3

1.贝叶斯判定准则

2.多元正态分布参数的极大似然估计

3.朴素贝叶斯分类器

# 贝叶斯判定准则

Bayes Decision Rule



贝叶斯判定准则:

为最小化总体风险，只需在每个样本上选择那个能使条件风险 $R(c|\boldsymbol{x})$ 最小的类别标记，即

$$h^*(\boldsymbol{x}) = \underset{c\in\mathcal{Y}}{\arg\min} R(c|\boldsymbol{x})$$

此时，$h^*$ 称为贝叶斯最优分类器。

# 贝叶斯判定准则

Bayes Decision Rule



已知条件风险 $R(c|\boldsymbol{x})$ 的计算公式为

$$R\left(c_i|\boldsymbol{x}\right)=\sum_{j=1}^{N}\lambda_{ij}P\left(c_j|\boldsymbol{x}\right)$$ 书上式7.1

若目标是最小化分类错误率，则误判损失 $\lambda_{ij}$ 对应为0/1损失，也即

$$\lambda_{ij}=\begin{cases}0, \text{ if } i=j\\1, \text{ otherwise}\end{cases}$$

那么条件风险 $R(c|\boldsymbol{x})$ 的计算公式可以进一步展开为

# 贝叶斯判定准则

Bayes Decision Rule

$$R(c_i|\boldsymbol{x}) = 1 \times P(c_1|\boldsymbol{x}) + ... + 1 \times P(c_{i-1}|\boldsymbol{x}) + 0 \times P(c_i|\boldsymbol{x}) + 1 \times P(c_{i+1}|\boldsymbol{x}) + ... + 1 \times P(c_N|\boldsymbol{x})$$

$$= P(c_1|\boldsymbol{x}) + ... + P(c_{i-1}|\boldsymbol{x}) + P(c_{i+1}|\boldsymbol{x}) + ... + P(c_N|\boldsymbol{x})$$

由于 $\sum_{j=1}^{N} P(c_j|\boldsymbol{x}) = 1$ ，所以

$$R(c_i|\boldsymbol{x}) = 1 - P(c_i|\boldsymbol{x})$$ 此即为式7.5

于是，最小化错误率的贝叶斯最优分类器为

$$h^*(\boldsymbol{x}) = \underset{c \in \mathcal{Y}}{\arg\min} R(c|\boldsymbol{x}) = \underset{c \in \mathcal{Y}}{\arg\min}\,(1 - P(c|\boldsymbol{x})) = \underset{c \in \mathcal{Y}}{\arg\max} P(c|\boldsymbol{x})$$

# 多元正态分布参数的极大似然估计

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters



已知对数似然函数为

$$LL(\boldsymbol{\theta}_c)=\sum_{\boldsymbol{x}\in D_c}\log P(\boldsymbol{x}|\boldsymbol{\theta}_c)$$ 书上式7.10

为了便于后续计算，我们令log的底数为e，则对数似然函数可化为

$$LL(\boldsymbol{\theta}_c)=\sum_{\boldsymbol{x}\in D_c}\ln P(\boldsymbol{x}|\boldsymbol{\theta}_c)$$

由于 $P(\boldsymbol{x}|\boldsymbol{\theta}_c)=P(\boldsymbol{x}|c)\sim\mathcal{N}\left(\boldsymbol{\mu}_c,\boldsymbol{\sigma}_c^2\right)$ ，那么

$$P(\boldsymbol{x}|\boldsymbol{\theta}_c)=\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_c|}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c)\right)$$

其中，d表示 $\boldsymbol{x}$ 的维数， $\boldsymbol{\Sigma}_c=\boldsymbol{\sigma}_c^2$ 为对称正定协方差矩阵，$|\boldsymbol{\Sigma}_c|$表示 $\boldsymbol{\Sigma}_c$ 的行列式，将上式代入对数似然函数可得

# 多元正态分布参数的极大似然估计



Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

$$LL(\boldsymbol{\theta}_c) = \sum_{\boldsymbol{x} \in D_c} \ln \left[ \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_c|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^{\mathrm{T}} \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_c)\right) \right]$$

令 $|D_c| = N$，则对数似然函数可化为

$$LL(\boldsymbol{\theta}_c) = \sum_{i=1}^{N} \ln \left[ \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_c|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^{\mathrm{T}} \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{x}_i-\boldsymbol{\mu}_c)\right) \right]$$

$$= \sum_{i=1}^{N} \ln \left[ \frac{1}{\sqrt{(2\pi)^d}} \cdot \frac{1}{\sqrt{|\boldsymbol{\Sigma}_c|}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^{\mathrm{T}} \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{x}_i-\boldsymbol{\mu}_c)\right) \right]$$

$$= \sum_{i=1}^{N} \left\{ \ln \frac{1}{\sqrt{(2\pi)^d}} + \ln \frac{1}{\sqrt{|\boldsymbol{\Sigma}_c|}} + \ln \left[ \exp\left(-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^{\mathrm{T}} \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{x}_i-\boldsymbol{\mu}_c)\right) \right] \right\}$$

# 多元正态分布参数的极大似然估计

deepshare.net
深度之眼

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

$$LL(\boldsymbol{\theta}_c)=\sum_{i=1}^{N}\left\{-\frac{d}{2}\ln(2\pi)-\frac{1}{2}\ln|\boldsymbol{\Sigma}_c|-\tfrac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)\right\}$$

$$=-\frac{Nd}{2}\ln(2\pi)-\frac{N}{2}\ln|\boldsymbol{\Sigma}_c|-\tfrac{1}{2}\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)$$

由于参数 $\boldsymbol{\theta}_c$ 的极大似然估计 $\hat{\boldsymbol{\theta}}_c$ 为

$$\hat{\boldsymbol{\theta}}_c=\underset{\boldsymbol{\theta}_c}{\arg\max}\,LL(\boldsymbol{\theta}_c)$$

所以接来下只需要求出使得对数似然函数 $LL(\boldsymbol{\theta}_c)$ 取到最大值的 $\hat{\boldsymbol{\mu}}_c$ 和 $\hat{\boldsymbol{\Sigma}}_c$ 也就求出了 $\hat{\boldsymbol{\theta}}_c$

# 多元正态分布参数的极大似然估计

deepshare.net 深度之眼

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

对 $LL(\boldsymbol{\theta}_c)$ 关于 $\boldsymbol{\mu}_c$ 求偏导

$$\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\mu}_c} = \frac{\partial}{\partial \boldsymbol{\mu}_c}\left[-\frac{Nd}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}_c| - \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)\right]$$

$$= \frac{\partial}{\partial \boldsymbol{\mu}_c}\left[-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)\right]$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\mu}_c}\left[(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)\right]$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\mu}_c}\left[(\boldsymbol{x}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)\right]$$

# 多元正态分布参数的极大似然估计

deepshare.net
深度之眼

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

$$\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\mu}_c} = -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\mu}_c}\left[(\boldsymbol{x}_i^T - \boldsymbol{\mu}_c^T)(\boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_i - \boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c)\right]$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\mu}_c}\left[\boldsymbol{x}_i^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_i - \boldsymbol{x}_i^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c - \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_i + \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c\right]$$

由于 $\boldsymbol{x}_i^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c$ 的计算结果为标量，所以

$$\boldsymbol{x}_i^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c = (\boldsymbol{x}_i^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c)^T = \boldsymbol{\mu}_c^T(\boldsymbol{\Sigma}_c^{-1})^T\boldsymbol{x}_i = \boldsymbol{\mu}_c^T(\boldsymbol{\Sigma}_c^T)^{-1}\boldsymbol{x}_i = \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_i$$

于是上式可以进一步化为

$$\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\mu}_c} = -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\mu}_c}\left[\boldsymbol{x}_i^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_i - 2\boldsymbol{x}_i^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c + \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c\right]$$

# 多元正态分布参数的极大似然估计

deepshare.net
深度之眼

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

由矩阵微分公式 $\frac{\partial \boldsymbol{a}^T \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}, \frac{\partial \boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x}}{\partial \boldsymbol{x}} = \left(\boldsymbol{B} + \boldsymbol{B}^T\right) \boldsymbol{x}$ 可得

$$
\begin{aligned}
\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\mu}_c} &= -\tfrac{1}{2} \sum_{i=1}^{N} \left[0 - (2\boldsymbol{x}_i^T \boldsymbol{\Sigma}_c^{-1})^T + \left(\boldsymbol{\Sigma}_c^{-1} + (\boldsymbol{\Sigma}_c^{-1})^T\right) \boldsymbol{\mu}_c\right] \\
&= -\tfrac{1}{2} \sum_{i=1}^{N} \left[-\left(2(\boldsymbol{\Sigma}_c^{-1})^T \boldsymbol{x}_i\right) + \left(\boldsymbol{\Sigma}_c^{-1} + (\boldsymbol{\Sigma}_c^{-1})^T\right) \boldsymbol{\mu}_c\right] \\
&= -\tfrac{1}{2} \sum_{i=1}^{N} \left[-(2\boldsymbol{\Sigma}_c^{-1} \boldsymbol{x}_i) + 2\boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c\right] \\
&= \sum_{i=1}^{N} \boldsymbol{\Sigma}_c^{-1} \boldsymbol{x}_i - N\boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c
\end{aligned}
$$

# 多元正态分布参数的极大似然估计



Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

令偏导数等于0可得

$$\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\mu}_c} = \sum_{i=1}^{N} \boldsymbol{\Sigma}_c^{-1} \boldsymbol{x}_i - N\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c = 0$$

$$N\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c = \sum_{i=1}^{N} \boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_i$$

$$N\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c = \boldsymbol{\Sigma}_c^{-1}\sum_{i=1}^{N} \boldsymbol{x}_i$$

$$N\boldsymbol{\mu}_c = \sum_{i=1}^{N} \boldsymbol{x}_i$$

$$\boldsymbol{\mu}_c = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i \Rightarrow \hat{\boldsymbol{\mu}}_c = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i$$

此即为式7.12

# 多元正态分布参数的极大似然估计

deepshare.net
深度之眼

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

对 $LL(\boldsymbol{\theta}_c)$ 关于 $\boldsymbol{\Sigma}_c$ 求偏导

$$\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\Sigma}_c} = \frac{\partial}{\partial \boldsymbol{\Sigma}_c}\left[-\frac{Nd}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}_c| - \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)\right]$$

$$= \frac{\partial}{\partial \boldsymbol{\Sigma}_c}\left[-\frac{N}{2}\ln|\boldsymbol{\Sigma}_c| - \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)\right]$$

$$= -\frac{N}{2}\cdot\frac{\partial}{\partial \boldsymbol{\Sigma}_c}\left[\ln|\boldsymbol{\Sigma}_c|\right] - \frac{1}{2}\sum_{i=1}^{N}\frac{\partial}{\partial \boldsymbol{\Sigma}_c}\left[(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)\right]$$

由矩阵微分公式 $\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\cdot(\mathbf{X}^{-1})^{T}, \frac{\partial \boldsymbol{a}^{T}\mathbf{X}^{-1}\boldsymbol{b}}{\partial \mathbf{X}} = -\mathbf{X}^{-T}\boldsymbol{a}\boldsymbol{b}^{T}\mathbf{X}^{-T}$ 可得

# 多元正态分布参数的极大似然估计

deepshare.net 深度之眼

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

$$\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\Sigma}_c} = -\frac{N}{2} \cdot \frac{1}{|\boldsymbol{\Sigma}_c|} \cdot |\boldsymbol{\Sigma}_c| \cdot (\boldsymbol{\Sigma}_c^{-1})^T - \frac{1}{2} \sum_{i=1}^{N} \left[ -\boldsymbol{\Sigma}_c^{-T} (\boldsymbol{x}_i - \boldsymbol{\mu}_c)(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-T} \right]$$

$$= -\frac{N}{2} \cdot (\boldsymbol{\Sigma}_c^{-1})^T - \frac{1}{2} \sum_{i=1}^{N} \left[ -\boldsymbol{\Sigma}_c^{-T} (\boldsymbol{x}_i - \boldsymbol{\mu}_c)(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-T} \right]$$

$$= -\frac{N}{2} \boldsymbol{\Sigma}_c^{-1} + \frac{1}{2} \sum_{i=1}^{N} \left[ \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_c)(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1} \right]$$

令偏导数等于0可得

$$\frac{\partial LL(\boldsymbol{\theta}_c)}{\partial \boldsymbol{\Sigma}_c} = -\frac{N}{2} \boldsymbol{\Sigma}_c^{-1} + \frac{1}{2} \sum_{i=1}^{N} \left[ \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_c)(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1} \right] = 0$$

# 多元正态分布参数的极大似然估计

deepshare.net 深度之眼

Maximum Likelihood Estimation of Multivariate Normal Distribution Parameters

$$-\frac{N}{2}\boldsymbol{\Sigma}_c^{-1} = -\frac{1}{2}\sum_{i=1}^{N}\left[\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^T\boldsymbol{\Sigma}_c^{-1}\right]$$

$$N\boldsymbol{\Sigma}_c^{-1} = \sum_{i=1}^{N}\left[\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^T\boldsymbol{\Sigma}_c^{-1}\right]$$

$$N\boldsymbol{\Sigma}_c^{-1} = \boldsymbol{\Sigma}_c^{-1}\left[\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^T\right]\boldsymbol{\Sigma}_c^{-1}$$

$$N = \boldsymbol{\Sigma}_c^{-1}\left[\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^T\right]$$

$$\boldsymbol{\Sigma}_c = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^T \Rightarrow \hat{\boldsymbol{\Sigma}}_c = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{\mu}_c)(\boldsymbol{x}_i-\boldsymbol{\mu}_c)^T$$

此即为式7.13

最值点证明：《多元正态分布参数的估计和数据的清洁与变换》（http://staff.ustc.edu.cn/~zwp/teach/MVA/Lec5_slides.pdf）

# 朴素贝叶斯分类器

Naive Bayes classifier



已知最小化分类错误率的贝叶斯最优分类器为

$$h^*(\boldsymbol{x}) = \underset{c \in \mathcal{Y}}{\arg\max} P(c|\boldsymbol{x})$$

又由贝叶斯定理可知

$$P(c|\boldsymbol{x}) = \frac{P(\boldsymbol{x}, c)}{P(\boldsymbol{x})} = \frac{P(c)P(\boldsymbol{x}|c)}{P(\boldsymbol{x})}$$

所以

$$h^*(\boldsymbol{x}) = \underset{c \in \mathcal{Y}}{\arg\max} \frac{P(c)P(\boldsymbol{x}|c)}{P(\boldsymbol{x})} = \underset{c \in \mathcal{Y}}{\arg\max} P(c)P(\boldsymbol{x}|c)$$

# 朴素贝叶斯分类器

Naive Bayes classifier



已知属性条件独立性假设为

$$P(\boldsymbol{x}|c) = P(x_1, x_2, ..., x_d|c) = \prod_{i=1}^{d} P(x_i|c)$$

所以

$$h^*(\boldsymbol{x}) = \arg\max_{c \in \mathcal{Y}} P(c) \prod_{i=1}^{d} P(x_i|c)$$

此即为朴素贝叶斯分类器的表达式

# 朴素贝叶斯分类器

Naive Bayes classifier

对于 $P(c)$

它表示的是样本空间中各类样本所占的比例，根据大数定律，当训练集包含充足的独立同分布样本时，$P(c)$ 可通过各类样本出现的频率来进行估计，也即

$$P(c) = \frac{|D_c|}{|D|}$$

其中，D表示训练集，$|D|$ 表示D中的样本个数，$D_c$表示训练集D中第c类样本组成的集合，$|D_c|$ 表示集合 $D_c$ 中的样本个数。

# 朴素贝叶斯分类器

Naive Bayes classifier

对于 $P(x_i|c)$

若样本的第i个属性 $x_i$ 取值为连续值:

我们假设该属性的取值服从正态分布，也即

$$P\left(x_i|c\right) \sim \mathcal{N}\left(\mu_{c,i}, \sigma_{c,i}^2\right) \Rightarrow P\left(x_i|c\right) = \frac{1}{\sqrt{2\pi}\sigma_{c,i}} \exp\left(-\frac{\left(x_i-\mu_{c,i}\right)^2}{2\sigma_{c,i}^2}\right)$$

其中正态分布的参数可以用极大似然估计法推得: $\mu_{c,i}$ 和 $\sigma_{c,i}^2$ 即为第c类样本在第i个属性上取值的均值和方差

# 朴素贝叶斯分类器

Naive Bayes classifier



对于 $P(x_i|c)$

若样本的第i个属性 $x_i$ 取值为离散值:

同样根据极大似然估计法，我们用其频率值作为其概率值的估计值，也即

$$P(x_i|c) = \frac{|D_{c,x_i}|}{|D_c|}$$

其中，$D_{c,x_i}$ 表示 $D_c$ 中在第i个属性上取值为 $x_i$ 的样本组成的集合。

# 朴素贝叶斯分类器

Naive Bayes classifier

deepshare.net
深度之眼

例：现将一枚6面骰子抛掷10次，抛掷出的点数分别为2、3、2、5、4、6、1、3、4、2，试基于此抛掷结果估计这枚骰子抛掷出各个点数的概率。

[解]：设这枚骰子抛掷出点数i的概率为 $P_i$，根据极大似然估计法可以写出似然函数为

$$L(\theta) = P_1 \times P_2^3 \times P_3^2 \times P_4^2 \times P_5 \times P_6$$

其对数似然函数即为

$$LL(\theta) = \ln L(\theta) = \ln(P_1 \times P_2^3 \times P_3^2 \times P_4^2 \times P_5 \times P_6)$$

$$= \ln P_1 + 3\ln P_2 + 2\ln P_3 + 2\ln P_4 + \ln P_5 + \ln P_6$$

# 朴素贝叶斯分类器

Naive Bayes classifier

由于 $P_i$之间满足如下约束

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 = 1$$

所以此时最大化对数似然函数属于带约束的最优化问题，也即

$$\max \quad LL(\theta) = \ln P_1 + 3\ln P_2 + 2\ln P_3 + 2\ln P_4 + \ln P_5 + \ln P_6$$

$$s.t. \quad P_1 + P_2 + P_3 + P_4 + P_5 + P_6 = 1$$

# 朴素贝叶斯分类器

Naive Bayes classifier



定理：对于一个优化问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) \leq 0 \quad (i=1,...,n) \\ & h_j(\boldsymbol{x}) = 0 \quad (j=1,...,m) \end{aligned}$$

若 $f(\boldsymbol{x}), g_i(\boldsymbol{x}), h_j(\boldsymbol{x})$ 一阶连续可微，并且$f(\boldsymbol{x}), g_i(\boldsymbol{x})$ 是凸函数，$h_j(\boldsymbol{x})$是线性函数，那么满足如下KKT条件的点 $\boldsymbol{x}^*$一定是优化问题的最优解。

$$\begin{cases} \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{n} \mu_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = 0 \\ h_j(\boldsymbol{x}^*) = 0 \\ g_i(\boldsymbol{x}^*) \leq 0 \\ \mu_i^* \geq 0 \\ \mu_i^* g_i(\boldsymbol{x}^*) = 0 \end{cases}$$

参考文献:王燕军，梁治安. 最优化基础理论与方法[M]. 复旦大学出版社，2011.

# 朴素贝叶斯分类器

Naive Bayes classifier



由拉格朗日乘子法可得拉格拉格朗日函数为

$$\mathcal{L}(\theta,\lambda)=\ln P_1+3\ln P_2+2\ln P_3+2\ln P_4+\ln P_5+\ln P_6+\lambda(P_1+P_2+P_3+P_4+P_5+P_6-1)$$

对拉格朗日函数 $\mathcal{L}(\theta)$ 分别关于 $P_i$ 求偏导，然后令其等于0可得

$$\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial P_1}=\frac{\partial}{\partial P_1}[\ln P_1+3\ln P_2+2\ln P_3+2\ln P_4+\ln P_5+\ln P_6+\lambda(P_1+P_2+P_3+P_4+P_5+P_6-1)]=0$$

$$=\frac{\partial}{\partial P_1}(\ln P_1+\lambda P_1)=0$$

$$=\frac{1}{P_1}+\lambda=0$$

$$\Rightarrow\lambda=-\frac{1}{P_1}$$

# 朴素贝叶斯分类器

Naive Bayes classifier



同理可求得:

$$\lambda = -\frac{1}{P_1} = -\frac{3}{P_2} = -\frac{2}{P_3} = -\frac{2}{P_4} = -\frac{1}{P_5} = -\frac{1}{P_6}$$

又因为

$$P_1 + P_2 + P_3 + P_4 + P_5 + P_6 = 1$$

所以最终解得

$$P_1 = \frac{1}{10}, P_2 = \frac{3}{10}, P_3 = \frac{2}{10}, P_4 = \frac{2}{10}, P_5 = \frac{1}{10}, P_6 = \frac{1}{10}$$

此时抛掷出各个点数的概率值与其频率值相等。

# 结 语

在这次课程中，我们学习了西瓜书

朴素贝叶斯的公式推导

那么在下次课程中，我们将会学习西瓜书

**神经网络的公式推导**