Lecture3
贝叶斯分类

贝叶斯分类

- 贝叶斯公式
- 最小错误率贝叶斯决策
- 最小风险贝叶斯决策
- 习题

- 给一个硬币，已知重量x，猜是5角还是1角
  - 后验概率 (posterior probability)

$$P(\omega_i|\mathrm{x})$$

  - 决策规则

$$x\begin{cases}\in\ \omega_1 & P(\omega_1|\mathrm{x})\ \geq P(\omega_2|\mathrm{x})\\ \in\ \omega_2 & P(\omega_1|\mathrm{x})\ < P(\omega_2|\mathrm{x})\end{cases}$$

  - 错误率

$$\mathrm{P}(\mathrm{error}|x) = 1 - \mathrm{P}(\omega_1|x) = \mathrm{P}(\omega_2|x) \quad (\text{如果决策 } x \in\ \omega_1)$$

  - 贝叶斯公式 (Bayes formula)

$$P(\omega_i|\mathrm{x}) = \frac{P(\omega_i,\mathrm{x})}{P(\mathrm{x})} = \frac{P(\mathrm{x}|\omega_i)P(\omega_i)}{P(\mathrm{x})}$$

Class Prior Probability

Likelihood

$$p(y|x)=\frac{p(x\,|\,y)\boxed{p(y)}}{\boxed{p(x)}}$$

Posterior Probability

Predictor Prior Probability

Likelihood

Class Prior Probability

$$p(y|x)=\frac{p(x\mid y)p(y)}{p(x)}$$

Posterior Probability

Predictor Prior Probability

似然函数 $p(x\mid y)$：根据样本出现的频率来估计这个特征的概率

后验概率 $p(y\mid x)$：事件x发生后，得到结果y的概率

概率vs似然

概率：已知硬币的参数，就可以去推测抛硬币的各种情况的可能性

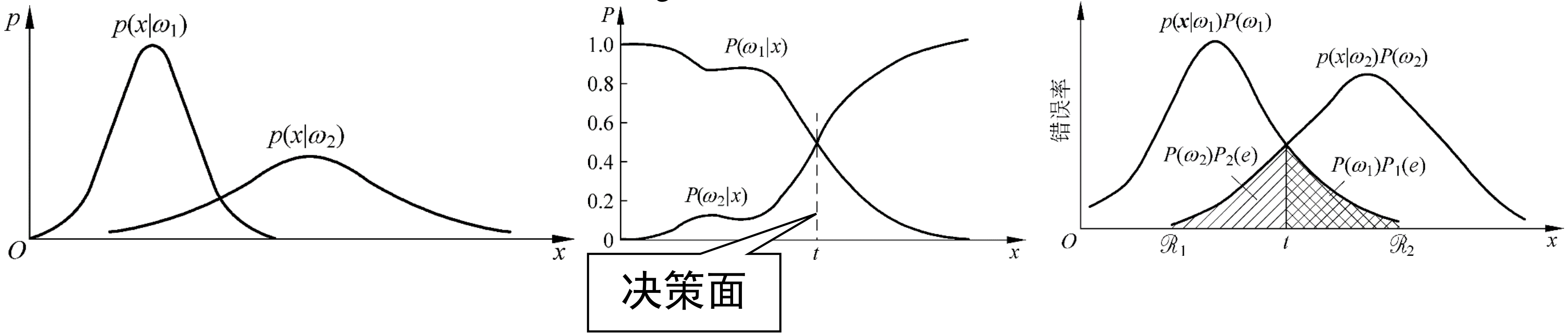
似然：我们对硬币的参数并不清楚，要通过抛硬币的情况去推测硬币的参数

最小错误率贝叶斯决策

样本x上错误的概率 $P(\text{error}|\text{x}) = \begin{cases} P(\omega_2|x) \text{ if } x \in \omega_1 \\ P(\omega_1|x) \text{ if } x \in \omega_2 \end{cases}$

所有服从同样分布的独立样本上错误概率的期望

$$P(\text{error}) = \int P(\text{error}|\text{x}) P(\text{x}) dx$$



决策规则　如果 $P(\omega_1|x) > P(\omega_2|x)$，则 $x \in \omega_1$；反之，则 $x \in \omega_2$

两类错误率

		真实类别	
		Positive	Negative
预测类别	Positive	TP	FP
	Negative	FN	TN

灵敏度 $S_n = \dfrac{TP}{TP+FN}$

特异度 $S_p = \dfrac{TN}{TN+FP}$

- P(Positive)和N（Negative）表示模型的判断结果
- T（True）和F（False）表示模型的判断结果是否正确

- FP：假正例（第一类错误）
- FN：假负例（第二类错误）
- TP：真正例
- TN：真负例

两类错误率

- 第一类错误率α：真实的阴性样本中被错误判断为阳性的比例
- 第二类错误率β：真实的阳性样本中被错误判断为阴性的比例

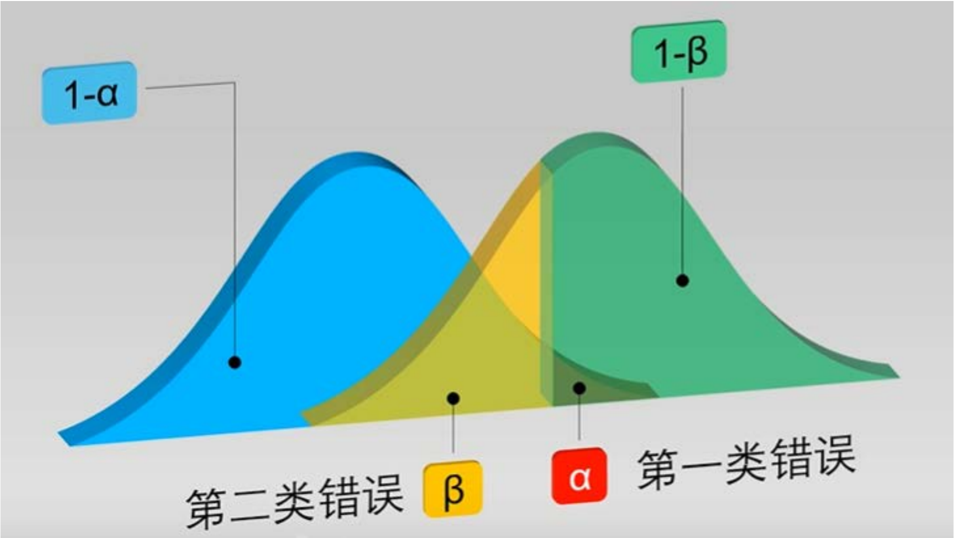
第一类错误率： $P_1(e)=\int_{R_2} p(x|\omega_1)dx$

第二类错误率： $P_2(e)=\int_{R_1} p(x|\omega_2)dx$

Neyman-Pearson决策

固定一类错误率、使另一类错误率尽可能小

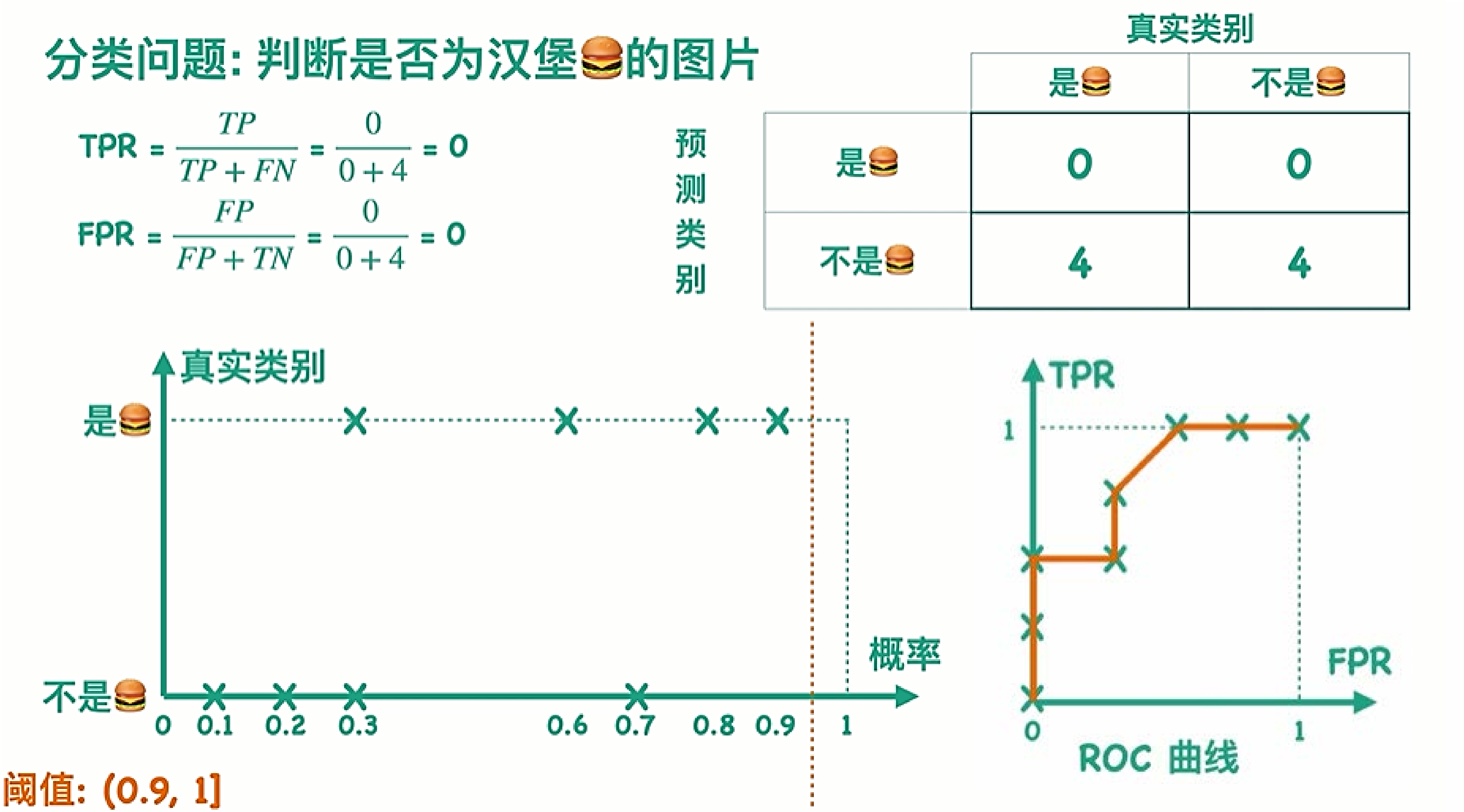
$$\min P_1(e)$$
$$s.t. P_2(e)-\varepsilon_0=0$$



ROC(receiver operating characteristic curve)曲线

对同一数据集在不同判定标准（阈值）下的结果

- ROC曲线能容易查出任意界限值对性能的识别能力
- ROC曲线越靠近左上角，试验的准确率越高



最小风险贝叶斯决策

条件风险（期望损失）： $R(c_i|x)=\sum_{j=1}^{N} k_{ij} P(c_j|x)$

1. 错误的分类会带来损失

把病人误诊为健康 ---> 风险代价大

把正常人误诊为病人 ---> 风险代价小

2. 不同的错误带来的损失可能不同，记作 $k_{ij} \begin{cases} i \rightarrow \text{判断类别} \\ j \rightarrow \text{真实类别} \end{cases}$

举例：c = 1, 2, 3, 4

$$R(c_1|x) = k_{1,1}P(c_1|x)+k_{1,2}P(c_2|x)+k_{1,3}P(c_3|x)+k_{1,4}P(c_4|x)$$

最小风险贝叶斯决策

判定准则：

总体风险： 最小化条件风险的期望

$$\mathrm{R(h)=E}_x[R(h(x)|x)] \longleftarrow \text{最小化}$$

最小化每一项 $R(c|x)$

$$\mathrm{h}^*(x) = \underset{c \in Y}{\arg\min}\, R(c|x) = \sum \mathrm{k_{ij}} P(c|x)$$

贝叶斯最优分类器

最小风险贝叶斯决策

0−1损失：$\mathrm{k}_{ij}=\begin{cases}0, & \text{if } \mathrm{i=j}\\ 1, & \text{otherwise}\end{cases}$

$$\mathrm{h}^*(x)=\underset{c\in Y}{\arg\min}\, R(c\mid x)$$

$$\mathrm{R}(c\mid x)=1-p(c\mid x)$$

$$\mathrm{R}(c|\mathrm{x}) = k_{1,1}P(c_1\mid x)+\mathrm{k}_{1,2}P(c_2\mid x)+k_{1,3}P(c_3\mid x)+k_{1,4}P(c_4\mid x)$$

$$\mathrm{h}^*(x)=\underset{c\in Y}{\arg\max}\, P(c\mid x)$$

练习

（1）假设在某个局部地区细胞识别中正常（ω_1）和异常（ω_2）两类的先验概率分别为：正常状态 $P(\omega_1)=0.9$ ，异常状态 $P(\omega_2)=0.1$ 。

现有一待识别的细胞，其观察值为 x，从类条件概率密度曲线上分别查得 $P(x|\omega_1)=0.2$，$P(x|\omega_2)=0.4$，试对该细胞 x 进行分类

（2）在（1）给出条件的基础上，利用如下决策表，按最小风险贝叶斯决策进行分类

决策	状态	
	ω_1	ω_2
α_1	0	6
α_2	1	0

解1：利用贝叶斯公式，分别计算 ω_1 及ω_2的后验概率

$$P(\omega_1 \mid x)=\frac{p(x \mid \omega_1)P(\omega_1)}{\sum_{j=1}^{2} p(x \mid \omega_j)P(\omega_j)}=\frac{0.2*0.9}{0.2*0.9+0.4*0.1}=0.818$$

$$P(\omega_2 \mid x)=1-P(\omega_1 \mid x)=0.182$$

根据贝叶斯决策规则式 如果$P(\omega_1|\mathrm{x})>P(\omega_2 \mid x)$,则$\mathrm{x}\in\omega_1$ ，因为

$$P(\omega_1 \mid x)=0.818>P(\omega_2 \mid x)=0.182$$

所以合理的决策是把x归类于正常状态（ω_1）

解2：已知条件为

$$P(\omega_1)=0.9,\quad P(\omega_2)=0.1 \qquad \lambda_{11}=0,\quad \lambda_{12}=6$$

$$P(x|\omega_1)=0.2,\quad P(x|\omega_2)=0.4 \qquad \lambda_{21}=1,\quad \lambda_{22}=0$$

根据解1的计算结果可知后验概率为： $P(\omega_1 \mid x)=0.818$，$P(\omega_2 \mid x)=0.182$

再按式 $R(\alpha_i|x)=\sum_{j=1}^{c}\lambda(\alpha_i \mid \omega_j)P(\omega_j \mid x)$ 计算条件风险：

$$R(\alpha_1 \mid x)=\sum_{j=1}^{2}\lambda_{1j}P(\omega_j \mid x)=\lambda_{12}P(\omega_2 \mid x)=1.092$$

$$R(\alpha_2 \mid x)=\lambda_{21}P(\omega_j \mid x)=0.818$$

$$R(\alpha_1 \mid x)>R(\alpha_2 \mid x)$$

即决策为ω_2的条件风险小于决策为ω_1的条件风险，因此采取决策行动 α_2，即判断待识别的细胞x 为ω_2类—异常细胞。

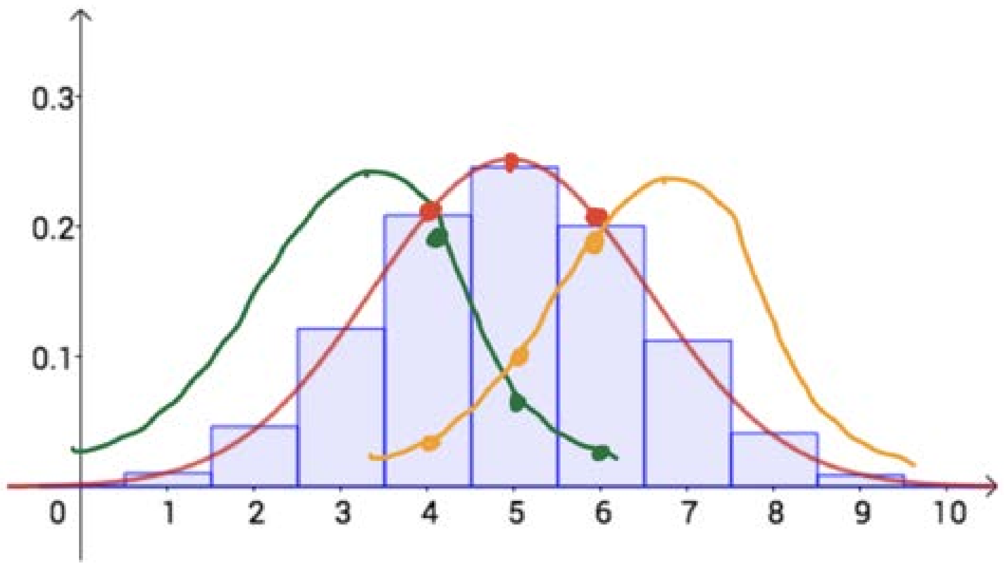
贝叶斯分类（参数估计）

- 最大似然估计
- 最大后验估计
- 贝叶斯估计

最大似然估计

原理：通过事实数据，猜测模型参数情况

- 相信试验/数据反映的客观规律
- 试验结果/数据既然已经发生，就说明该事件发生的可能性是很大的，是个大概率事件



最大似然估计求解——抛硬币问题

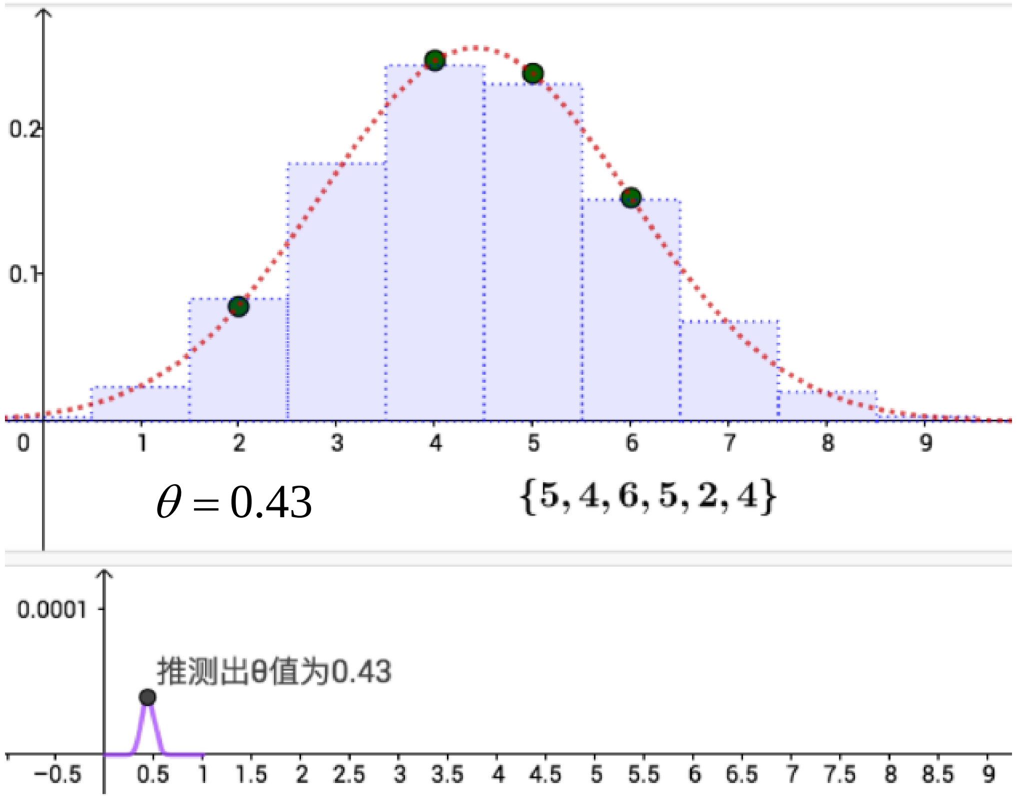
6次实验结果（每次抛10枚，统计“花”的次数）：$x_1, x_2, \ldots, x_6 = \{5, 4, 6, 5, 2, 4\}$

$$L(\theta) = [C_{10}^{5}\theta^{5}(1-\theta)^{5}][C_{10}^{4}\theta^{4}(1-\theta)^{6}]\ldots$$

最大似然函数步骤：

- 写似然函数 $L(\theta)$
- 取对数 $\ln L(\theta)$
- 求偏导 $\dfrac{\partial \ln L(\theta)}{\partial \theta} = 0$

$$\hat{\theta} \longrightarrow \max \mathrm{L}(\theta)$$



估计抛一次硬币“花”的概率为0.43

正态分布下的最大似然估计

概率密度函数 $$f(x)=\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(x-\mu)^2}{2\delta^2}}$$

写似然函数 $$L(\mu,\delta^2)=\prod_{i=1}^{N}f(x_i\mid\theta)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(x_i-\mu)^2}{2\delta^2}}=(2\pi\delta^2)^{-\frac{n}{2}}e^{\frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

取对数 $$\ln L(u,\sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-u)^2$$

求偏导 $$\begin{cases}\dfrac{\partial\ln L(u,\sigma^2)}{\partial u}=\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-u)=0\\[2ex]\dfrac{\partial\ln L(u,\sigma^2)}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-u)^2=0\end{cases}\longrightarrow\boxed{\begin{cases}u^*=\overline{x}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i\\[2ex]\sigma^{*2}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2\end{cases}}$$

最大后验估计MAP

最大后验估计与最大似然估计相似

不同点：在于估计θ的函数中允许加入一个先验 $p(\theta)$，此时不是要求似然函数最大，而是要求由贝叶斯公式计算出的整个后验概率最大

$$\hat{\theta}_{MAP} = \arg\max \frac{p(X\mid\theta)p(\theta)}{p(X)}$$

$$= \arg\max p(X\mid\theta)p(\theta)$$

$$= \arg\max\{L(\theta\mid X) + \log p(\theta)\}$$

$$= \arg\max\{\sum_{x\in X} \mathrm{p}(\theta\mid X) + \log p(\theta)\}$$

贝叶斯估计（Bayesian Estimation）

- 基本思想
  - ✓ 把参数估计看成贝叶斯决策问题
  - ✓ 待估计参数看成具有先验分布$p(\theta)$的随机变量
  - ✓ 其取值和样本集有关 $\chi=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N\}$
  - ✓ 最优条件是最小化错误率或者风险，损失函数为$\lambda\left(\widehat{\boldsymbol{\theta}},\theta\right)$

- 用$\widehat{\boldsymbol{\theta}}$来估计时的总期望风险为

  样本取值空间E^d,参数取值空间为Θ

$$R=\int_{E^d}\int_{\Theta}\lambda\left(\widehat{\boldsymbol{\theta}},\theta\right)\ p(\theta|\boldsymbol{x})\,p(\boldsymbol{x})d\theta d\boldsymbol{x}$$

贝叶斯估计

$p(\theta)$ 是待估计参数 θ 的先验概率，θ 取值与样本集 $D=\{x_1,x_2,\ldots x_N\}$ 有关，$\lambda(\hat{\theta},\theta)$ 是 $\hat{\theta}$ 作为θ 的估计量时的损失函数。

定义条件风险函数为

$$R(\hat{\theta}\mid x)=\int\lambda(\hat{\theta},\theta)p(\theta\mid x)d\theta$$

则有：

$$R=\int_{E^d}R(\hat{\theta}|x)p(x)dx$$

又因 $R(\hat{\theta}\mid x)$ 非负，则由贝叶斯决策知，求R最小即求 $R(\hat{\theta}\mid x)$ 最小，即：

$$\theta^*=\arg\min R(\hat{\theta}\mid x)$$

可得最优估计： $\theta^*=\int\theta p(\theta\mid x)d\theta$

贝叶斯估计步骤

贝叶斯估计的基本步骤(基于平方误差损失函数)：

1、确定参数 θ 的先验分布 $P(\theta)$

2、由样本集 $D=\{x_1,x_2,...,x_n\}$ 求出样本联合分布 $P(D|\theta)$ ，

其中： $$P(D|\theta)=\prod_{n=1}^{N}P(x_n|\theta)$$

3、利用贝叶斯公式, 求后验分布 $$P(\theta|D)=\frac{P(D|\theta)P(\theta)}{\int_\theta P(D|\theta)P(\theta)d\theta}$$

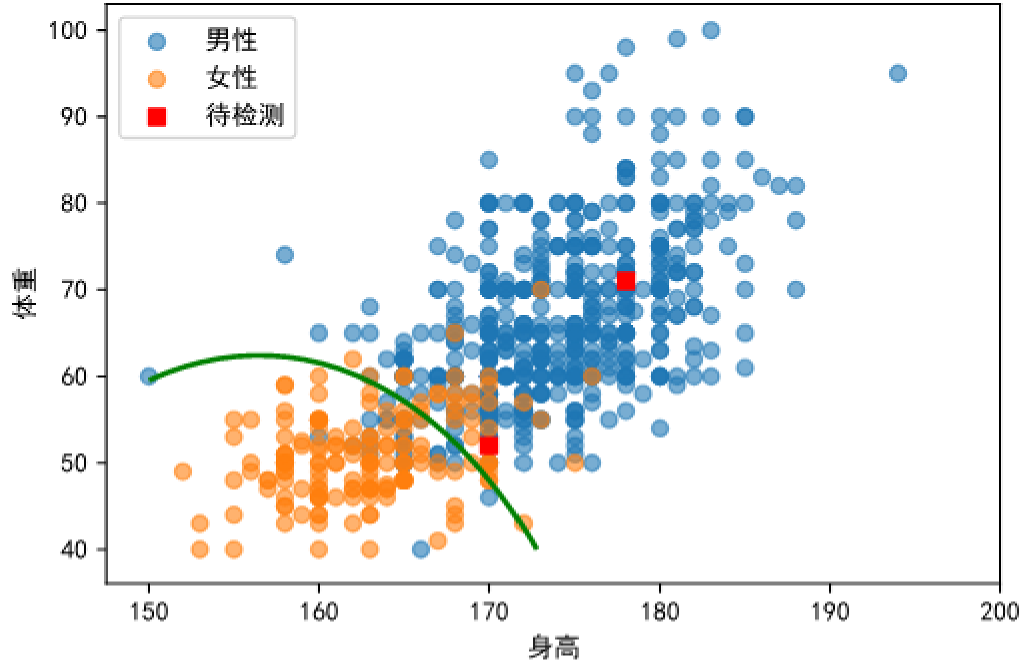
4、求出贝叶斯估计值 $$\hat{\theta}=\int_\theta \theta P(\theta|D)d\theta$$

作业二：利用贝叶斯分类器实现基于身高体重的性别分类

- 推荐编程环境：Anaconda+Jupyter notebook+pytorch

  安装教程： 点这

- 要求可视化决策面以及分类结果



Sources

- https://www.youtube.com/watch?v=60pqgfT5tZM
  Text Classification Using Naive Bayes | Simplilearn 20mins

- https://www.youtube.com/watch?v=O2L2Uv9pdDA
  Naive Bayes, Clearly Explained!!! 15mins

- https://www.youtube.com/watch?v=9wCnvr7Xw4E
  Bayes' Theorem, Clearly Explained!!!! 14mins

- https://www.youtube.com/watch?v=XQoLVl31ZfQ
  Bayes' Theorem - The Simplest Case 5mins