SHENZHEN UNIVERSITY 深圳大学

# 模式识别

# Pattern Recognition

——与计算机视觉

and Computer Vision


樊超

大数据系统计算技术国家工程实验室@深圳大学

**E-mail：** chaofan996@szu.edu.cn

# 贝叶斯决策

- 贝叶斯法则
- 朴素贝叶斯决策
- 最小错误率贝叶斯决策
- 最小损失贝叶斯决策
- 评价指标与ROC曲线
- 贝叶斯参数估计：单维Gaussian估计（$\mu$未知）
- 贝叶斯参数估计：多维Gaussian估计（$\mu$未知）
- 关于Gaussian分布

# 贝叶斯法则

$$\underset{\text{后验（Posterior）}}{P(Y|X)} = \frac{\underset{\text{似然（Likelihood）}}{P(X|Y)}\ \underset{\text{先验（Priors）}}{P(Y)}}{P(X)}$$

Rev'd Thomas Bayes (1702–1761)

- 贝叶斯法则：如何根据已知 **数据** $P(X|Y)$ 与 **先验** $P(Y)$ 推断 **后验** $P(Y|X)$

# 贝叶斯法则——一个简单例子

$$\underbrace{P(Y|X)}_{\substack{\text{后验}\\ \text{（Posterior）}}} = \frac{\overbrace{P(X|Y)}^{\substack{\text{似然}\\ \text{（Likelihood）}}} \; \overbrace{P(Y)}^{\substack{\text{先验}\\ \text{（Priors）}}}}{P(X)}$$

- 问题：教室里有50位同学，判断下位进门的同学是 **男孩** 还是 **女孩**?
- 令随机变量 $X$ 表示同学， $Y$ 表示性别
- 则先验分布 $P(Y)$ 表示教室里的性别分布，如$P(Y=\text{男})=\frac{45}{50}$， $P(Y=\text{女})=\frac{5}{50}$
- $P(X|Y)$ 表示观测数据：这一群是**男孩**，另一群是 **女孩**； $P(Y|X)$ 表示任意新进同学的性别分布

# 贝叶斯法则——另一个简单例子

例子：流感检测

假设某个城市的人口中，只有 **1%** 患有流感病毒。

现有一种检测方法，已知：

（1）**如果某人确实患病**，检测为阳性的概率为**90%**；

（2）**如果某人没有患病**，检测仍然有**5%**可能性误报为阳性。

现在，已知某人的检测结果为阳性，那么他真正患病的概率有多少？

# 贝叶斯法则——另一个简单例子

例子：流感检测

假设某个城市的人口中，只有 **1%** 患有流感病毒。

现有一种检测方法，已知：

（1）**如果某人确实患病**，检测为阳性的概率为**90%**；

（2）**如果某人没有患病**，检测仍然有**5%**可能性误报为阳性。

现在，已知某人的检测结果为阳性，那么他真正患病的概率有多少？

$X = x^+$　　　　$Y = y^+$

令随机变量 $X \in \{x^+, x^-\}$ 表示观测结果， $x^+$表示阳性， $x^-$表示阴性

令随机变量 $Y \in \{y^+, y^-\}$ 表示真实结果， $y^+$表示阳性， $y^-$表示阴性

# 贝叶斯法则——另一个简单例子

例子：流感检测

假设某个城市的人口中，只有 **1%** 患有流感病毒。

现有一种检测方法，已知：

（1）**如果某人确实患病**，检测为阳性的概率为**90%**；

（2）**如果某人没有患病**，检测仍然有**5%**可能性误报为阳性。

现在，已知某人的$\boxed{\text{检测结果为阳性}}$，那么他$\boxed{\text{真正患病}}$的概率有多少？

$$\underbrace{X = x^+ \qquad\qquad Y = y^+}_{\text{推断 } P(y^+|x^+)\text{？}}$$

$$P(Y|X) = \frac{P(X|Y)\ P(Y)}{P(X)}$$

# 贝叶斯法则——另一个简单例子

例子：流感检测

假设某个城市的人口中，只有 **1%** 患有流感病毒。 $\Rightarrow$ $P(y^+) = 0.01$

现有一种检测方法，已知：

（1）**如果某人确实患病**，检测为阳性的概率为**90%**； $\Rightarrow$ $P(x^+|y^+) = 0.9$

（2）**如果某人没有患病**，检测仍然有**5%**可能性误报为阳性。 $\Rightarrow$ $P(x^+|y^-) = 0.05$

现在，已知某人的检测结果为阳性，那么他真正患病的概率有多少？

$$\underbrace{X = x^+ \qquad\qquad Y = y^+}_{\text{推断 } P(y^+|x^+)\text{？}}$$

$$P(Y|X) = \frac{P(X|Y)\ P(Y)}{P(X)}$$

# 贝叶斯法则——另一个简单例子

已知：

（1）城市中任意一人，真实患病的概率 $P(y^+) = 0.01$

（2）若真实患病，被检测出来患病的概率 $P(x^+|y^+) = 0.9$；

（3）若没有患病，被检测出来患病的概率$P(x^+|y^-) = 0.05$；

求：检测阳性，真实阳性的概率 $P(y^+|x^+) = \frac{P(x^+|y^+)P(y^+)}{\boldsymbol{P}(x^+)}$

$$P(x^+) = P(x^+|y^+)P(y^+) + P(x^+|y^-)\,P(y^-) = 0.9 * 0.01 + 0.05 * (1 - 0.05) = 0.0585$$

$$P(y^+|x^+) = \frac{P(x^+|y^+)P(y^+)}{P(x^+)} = \frac{0.9 * 0.01}{0.0585} \approx 0.154$$

$$P(Y|\mathrm{X}) = \frac{P(X|Y)\;P(Y)}{P(X)}$$

# 朴素贝叶斯决策

已知：

（1）城市中任意一人，真实患病的概率$P(y^+) = 0.01$

（2）若真实患病，被检测出来患病的概率$P(x^+|y^+) = 0.9$；

（3）若没有患病，被检测出来患病的概率$P(x^+|y^-) = 0.05$；

决策规则：若$P(y^+|x^+) > P(y^-|x^+)$， 则😭 （检测阳性，推断阳性）

否则，$P(y^+|x^+) \leq P(y^-|x^+)$， 则😊（检测阳性，推断阴性）

$$P(y^+|x^+) = \frac{P(x^+|y^+)P(y^+)}{P(x^+)} \qquad P(y^-|x^+) = \frac{P(x^+|y^-)P(y^-)}{P(x^+)}$$

# 朴素贝叶斯决策

😭 $P(y^+|x^+)$ $\quad VS. \quad$ $P(y^-|x^+)$ 😊

$$\Downarrow \qquad\qquad \Downarrow$$

$$\frac{P(x^+|y^+)P(y^+)}{\cancel{P(x^+)}} \quad VS. \quad \frac{P(x^+|y^-)P(y^-)}{\cancel{P(x^+)}}$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$P(x^+|y^+)P(y^+) \quad VS. \quad P(x^+|y^-)P(y^-)$$

$$\Rightarrow \quad \underbrace{P(Y|X)}_{\text{后验（Posterior）}} \propto \underbrace{P(X|Y)}_{\text{似然（Likelihood）}} \underbrace{P(Y)}_{\text{先验（Priors）}}$$

$$\text{推断} = \begin{cases} \text{😭 阳性，若} P(x^+|y^+)P(y^+) > P(x^+|y^-)P(y^-) \\ \text{😊 阴性，若} P(x^+|y^+)P(y^+) \le P(x^+|y^-)P(y^-) \end{cases}$$

# 朴素贝叶斯决策——一个简单的例子

已知：

（1）城市中任意一人，真实患病的概率$P(y^+) = 0.01$

（2）若真实患病，被检测出来患病的概率$P(x^+|y^+) = 0.9$；

（3）若没有患病，被检测出来患病的概率$P(x^+|y^-) = 0.05$；

😭 $P(x^+|y^+)P(y^+)$ $\quad VS. \quad$ $P(x^+|y^-)P(y^-)$ 😊

$$0.9 \times 0.01 = 0.009 \quad < \quad 0.05 \times (1 - 0.01) = 0.0495$$

推断：检出阳性，但问题不大😊

后验（Posterior） 似然（Likelihood） 先验（Priors）

$$P(Y|X) \propto P(X|Y)\ P(Y)$$

# 朴素贝叶斯决策——一个简单的例子

已知：

（1）城市中任意一人，真实患病的概率$P(y^+) = 0.1$

（2）若真实患病，被检测出来患病的概率$P(x^+|y^+) = 0.9$；

（3）若没有患病，被检测出来患病的概率$P(x^+|y^-) = 0.05$；

😭 $P(x^+|y^+)P(y^+)$ $VS.$ $P(x^+|y^-)P(y^-)$ 😊

$$0.9 \times 0.1 = 0.09 \quad > \quad 0.05 \times (1 - 0.1) = 0.045$$

推断：检出阳性，问题很大！ 😭

后验（Posterior） 似然（Likelihood） 先验（Priors）

$$P(Y|X) \propto P(X|Y)\ P(Y)$$

# 贝叶斯法则

## 先验影响推断

后验（Posterior）　似然（Likelihood）　先验（Priors）

$$P(Y|X) = \frac{P(X|Y)\ P(Y)}{P(X)}$$

➢ 什么是先验（ Priors ）？ ⟶ 期望/想象/信念

- 先验 可以很主观：觉得今天运气特别好，即便知道买彩票是捐赠行为，还是作决定去买
- 先验 可以很客观：某个彩票站开了大奖，即便知道买彩票是捐赠行为，还是作决定去买
- 先验 可以是模型：在DeepSeek的指导下，即便知道买彩票是捐赠行为，还是作决定去买
- 先验 可以是经验：很久没开某个号码了，即便知道买彩票是捐赠行为，还是作决定去买

Rev'd Thomas Bayes (1702–1761)

# 朴素贝叶斯决策——一个简单的例子

已知：

（1）城市中任意一人，真实患病的概率$P(y^+) = 0.01$

（2）若真实患病，被检测出来患病的概率$P(x^+|y^+) = 0.999$；

（3）若没有患病，被检测出来患病的概率$P(x^+|y^-) = 0.01$；

😭 $P(x^+|y^+)P(y^+)$ VS. $P(x^+|y^-)P(y^-)$ 😊

$$0.999 \times 0.01 = 0.00999 > 0.01 \times (1 - 0.01) = 0.0099$$

推断：检出阳性，问题很大！😭

后验（Posterior） 似然（Likelihood）先验（Priors）

$$P(Y|X) \propto P(X|Y)\ P(Y)$$

# 贝叶斯法则

**似然影响推断**

$$\underset{\text{后验（Posterior）}}{P(Y|X)} = \frac{\overset{\text{似然（Likelihood）}}{P(X|Y)}\ \overset{\text{先验（Priors）}}{P(Y)}}{P(X)}$$

- 什么是先验（ Priors ）？ $\longrightarrow$ 期望/想象/信念
- 什么是似然（Likelihood）？ $\longrightarrow$ 观测/证据/度量
- 什么是$P(X) = \int P(x|y)P(y)dy$ ? $\rightarrow$ 证据，模型簇 $Y$ 对数据 $X$ 的拟合
- 什么是$P(Y|X)$? $\longrightarrow$ 推断/不确定性

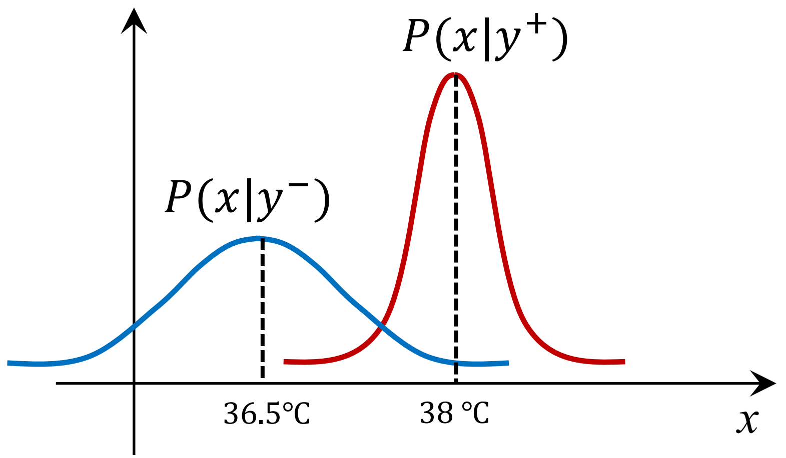
Rev'd Thomas Bayes (1702–1761)

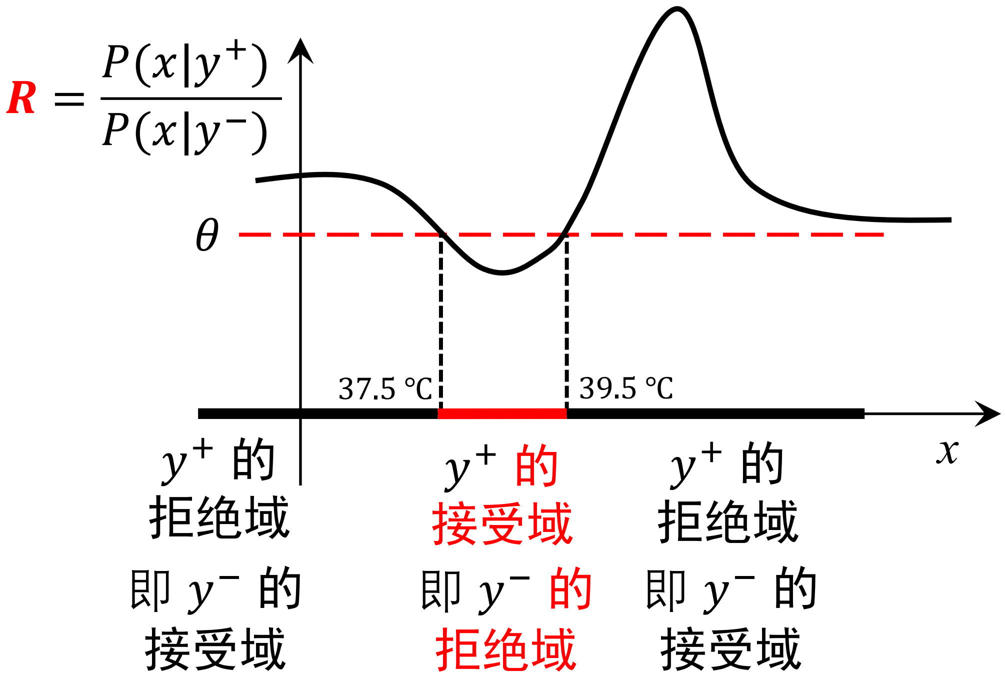
# 贝叶斯决策——最小错误率决策

贝叶斯最小错误率决策：$\max_{y} P(Y|X)$

$$
决策规则 = \begin{cases} 😭\ 推断阳性，若\ \dfrac{P(x^+|y^+)}{P(x^+|y^-)} > \dfrac{P(y^+)}{P(y^-)} \\ 😊\ 推断阴性，若\ \underbrace{\dfrac{P(x^+|y^+)}{P(x^+|y^-)}}_{\substack{似然比\ R \\ \text{（Likelihood Ratio）}}} \leq \dfrac{P(y^+)}{P(y^-)} \end{cases}
$$

# 贝叶斯决策——最小错误率决策




此前随机变量 $X = \{x^+, x^-\}$ 表示体温高低，此处令 $X = x$ 表示温度数值，即考虑连续情况

# 贝叶斯决策——最小损失决策

考虑多分类问题，模型$f_\theta(x)$，样本 $x$ 的真实标签为第 $y=j$ 类，被误判为第 $f_\theta(x)=i$ 类损失为：

$$\lambda_{ij}=\lambda(f_\theta(x)=i|y=j)$$

一般而言，当$i=j$ 时，$\lambda_{ij}=0$

- 损失函数： $R(f_\theta(x)|x)=\sum_i \lambda(f_\theta(x)|y=i)\, p(y=i|x)$
- 优化目标：$\min\limits_\theta \mathbb{E}(f_\theta(x)|x)$

特别地，当 $\lambda_{ij}=1$ （$i\neq j$）时，最小损失决策与最小错误率决策等效。

- $\lambda_{ij}$度量了不同决策的后果，如雷达检测敌机、传染病检测等场景需要提高系统灵敏度。

# 贝叶斯决策——评价指标

| 决策 \ 真实状态 | 阳性 | 阴性 |
|---|---|---|
| 阳性 | 真阳（TP） | 假阳（FP） |
| 阴性 | 假阴（FN） | 真阴（TN） |

T for True, F for False
P for Positive, N for Negative

两类错误率及评价指标

两类：阳性 (positive)、阴性 (negative)

第一类错误率 (Type-I error rate) $\frac{FP}{FP+TN}$ 误报/虚警

第二类错误率 (Type-II error rate) $\frac{FN}{FN+TP}$ 漏报

灵敏度 (sensitivity) $S_n = \frac{TP}{TP+FN}$

特异度 (specificity) $S_p = \frac{TN}{TN+FP}$

# 贝叶斯决策——评价指标

| 决策 \ 真实状态 | 阳性 | 阴性 |
|---|---|---|
| 阳性 | 真阳（TP） | 假阳（FP） |
| 阴性 | 假阴（FN） | 真阴（TN） |

T for True, F for False
P for Positive, N for Negative

其他常用评价指标

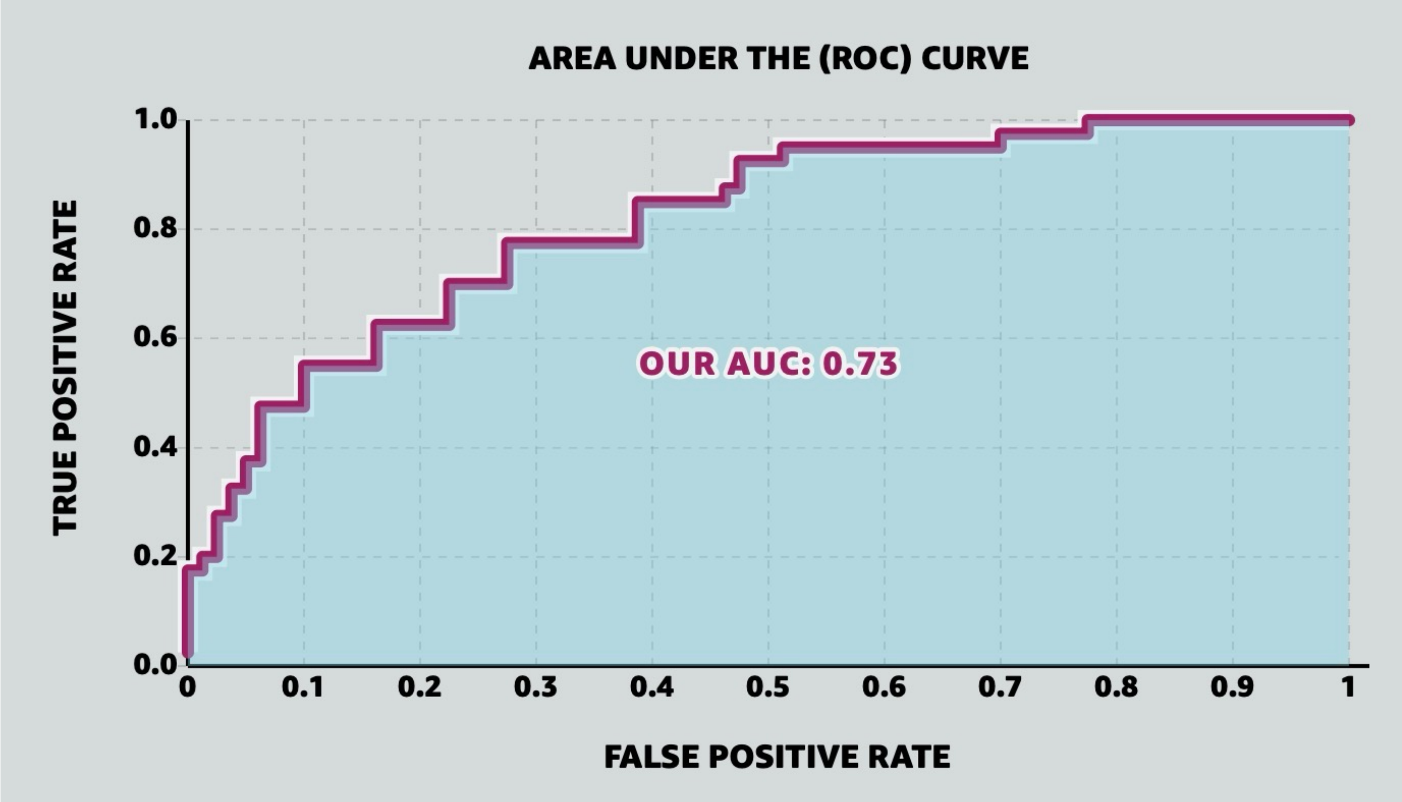
正确率 (accuracy) $\mathrm{Acc} = \dfrac{\mathrm{TP}+TN}{\mathrm{TP}+TN+FP+FN}$

召回率 (recall) $\mathrm{Rec} = \dfrac{TP}{TP+FN}$ 灵敏度

准确率 (precision) $\mathrm{Pre} = \dfrac{TP}{TP+FP}$

F度量 (F-measure) $F = \dfrac{2Rec\cdot Pre}{Rec+\mathrm{Pre}}$

# 贝叶斯决策——评价指标




参考博客：https://mlu-explain.github.io/roc-auc/

# 贝叶斯决策——推广

实际情况比上述“流感问题”复杂得多：

- 多变量
- 多分类
- 多目标
- 连续变量
- 复杂分布
- 带噪数据

$\xrightarrow{\text{参数化}}$

$$P(Y|X) = \frac{P(X|Y)\ P(Y)}{P(X)}$$

$$Y = f_{\theta}(X) \quad \downarrow$$

$$p(\theta|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\theta)\ p(\theta)}{p(\boldsymbol{x})}$$

# 贝叶斯参数估计——单维Gaussian估计（ $\mu$ 未知）

假设：

（1）$p(x|\mu) \sim N(\mu, \sigma^2)$ 且 $p(\mu) \sim N(\mu_0, \sigma_0^2)$，仅$\mu$未知，要求推断$\mu$

（2）数据集 $\mathcal{X} = \{x_1, \ldots, x_n\}$，采样独立

$$\max_{\mu} p(\mu|\mathcal{X}) = \frac{p(\mathcal{X}|\mu)\ p(\mu)}{p(\mathcal{X})} = \alpha \prod_{k=1}^{n} p(x_k|\mu) p(\mu)$$

$$p(\mu|\mathcal{X}) = \alpha \overbrace{\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x_k - \mu}{\sigma}\right)^2\right]}^{p(x_k|\mu)} \overbrace{\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right]}^{p(\mu)}$$

# 贝叶斯估计——单维Gaussian估计（$\mu$未知）

$$p(\mu|\mathcal{X}) = \alpha \underbrace{\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x_k-\mu}{\sigma}\right)^2\right]}_{p(x_k|\mu)} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right]}_{p(\mu)}$$

$$= \alpha' \exp\left[-\frac{1}{2}\left[\sum_{k=1}^{n}\left(\frac{\mu-x_k}{\sigma}\right)^2 + \left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right]\right]$$

$$= \alpha'' \exp\left[-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{1}{\sigma^2}\sum_{k=1}^{n} x_k + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right]$$

# 贝叶斯估计——单维Gaussian估计（$\mu$未知）

$$p(\mu|\mathcal{X}) = \alpha'' \exp\left[-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{1}{\sigma^2}\sum_{k=1}^{n} x_k + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right]$$

$$p(\mu|\mathcal{X}) = \frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right]$$

其中: $\sigma_n^2 = \dfrac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}$

$$\mu_n = \left(\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\right)\bar{x}_n + \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0$$

$$\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

- 当$\mu=\mu_n$ 时，后验概率 $p(\mu|\mathcal{X})$ 取得极大值
- $\sigma_n^2$度量了估计的不确定性

# 贝叶斯估计——单维Gaussian估计（$\mu$未知）

$$\mu_n = \left(\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right)\bar{x}_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\mu_0$$

- $\mu_n$ 是 $\bar{x}_n$ 与 $\mu_0$ 的线性组合
- 若 $\sigma_0 \neq 0$，当 $n \to \infty$（更多样本）， $\mu_n \to \bar{x}_n$
- 若 $\sigma_0 = 0$，则 $\mu_n = \mu_0$
- 若 $\sigma_0 \gg \sigma$，则 $\mu_n = \bar{x}_n$

# 贝叶斯估计——单维Gaussian估计（$\mu$未知）

$$\sigma_n^2 = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}$$

- 当 $n$ 持续增大（样本持续增多）， $\sigma_n^2$ 将 逼近 $\frac{\sigma^2}{n}$

  （更多的训练样本可以降低对$\mu$估计的不确定性）
- 当 $n$ 持续增大（样本持续增多），后验分布$p(\mu|\mathcal{X})$ 变得尖锐

  （方差变小）

# 贝叶斯估计——多维Gaussian估计（$\mu$未知）

假设：

（1）$p(\boldsymbol{x}|\boldsymbol{\mu}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 且 $p(\boldsymbol{\mu}) \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$，仅 $\boldsymbol{\mu}$ 未知，要求推断 $\boldsymbol{\mu}$

（2）数据集 $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$，采样独立

$$\max_{\boldsymbol{\mu}} p(\boldsymbol{\mu}|\mathcal{X}) = \frac{p(\mathcal{X}|\boldsymbol{\mu})\ p(\boldsymbol{\mu})}{p(\boldsymbol{x})} = \alpha \prod_{k=1}^{n} p(\boldsymbol{x}_k|\boldsymbol{\mu})p(\boldsymbol{\mu})$$

$$p(\boldsymbol{\mu}|\mathcal{X}) = \alpha \exp[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)^t \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_n)]$$

其中:

$$\boldsymbol{\mu}_n = \boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_0 + \frac{1}{n}\boldsymbol{\Sigma})^{-1}\overline{\boldsymbol{x}}_n + \frac{1}{n}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}_0 + \frac{1}{n}\boldsymbol{\Sigma})^{-1}\boldsymbol{\mu}_0 \qquad \overline{\boldsymbol{x}}_n = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{x}_k$$

$$\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_0 + \frac{1}{n}\boldsymbol{\Sigma})^{-1}\frac{1}{n}\boldsymbol{\Sigma}$$

# 关于Gaussian分布

许多同学知道中心极限定理（Central Limit Theorem，CLT）：

令 $\{x_1, \ldots, x_n\}$ 是从一个总体抽取的独立同分布的随机变量，具有期望值 $\mu$ 和方差 $\sigma^2$，

当样本量 $n$ 足够大时，样本均值 $\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$ 的分布会收敛于Gaussian分布：

$$\frac{\bar{x}_n - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

现实意义：

- 许多特征变量分布往往是非Gaussian的，但数据聚合（如均值、加权）后，往往接近Gaussian分布
- 测试误差往往由多个随机误差叠加，中心极限定理解释了这些误差常常服从Gaussian分布

# 关于Gaussian分布

还有一个很好的性质， Gaussian分布是最小假设（最大熵假设）

- 熵（Entropy）度量了分布的不确定性：
- 离散情况

$$H(p_1, \dots, p_n) = -\sum_i p_i \log(p_i)$$

- 当$p_i = \frac{1}{n}$时，熵最大（以n=2验算一下？）
- 连续情况：$h\big(p(x)\big) = -\int p(x) \log\big(p(x)\big)\, dx$
- 对Gaussian分布$g(x) \sim N(\mu, \sigma^2)$: $h(g(x)) = \frac{1}{2}\log(2\pi e\sigma^2)$

# 关于Gaussian分布

还有一个很好的性质， Gaussian分布是最小假设（最大熵假设）

- KL散度（KL Divergence）是对分布之间一致性的度量
- 离散情况：$D_{\mathrm{KL}}(P||Q) = \sum_i \log(\frac{p_i}{q_i})p_i$
- 连续情况： $D_{\mathrm{KL}}(p||q) = \int \log\left(\frac{p(x)}{q(x)}\right)p(x)dx$
- KL散度的性质：
  - 非对称，即$D_{\mathrm{KL}}(p||q) \neq D_{\mathrm{KL}}(q||p)$（JS散度具有对称性： $D_{\mathrm{JS}}(p||q)) = \frac{D_{\mathrm{KL}}(p||q)+D_{\mathrm{KL}}(q||p)}{2}$）
  - $D_{\mathrm{KL}}(p||q) \geq 0$
  - $D_{\mathrm{KL}}(p||q) = 0$ 当且仅当二者分布相同时成立

# 关于Gaussian分布

还有一个很好的性质， Gaussian分布是最小假设（最大熵假设）

假设： $p(x)$ 是一个均值为 $\mu$，方差为 $\sigma^2$ 的任意概率分布，

证明：在所有具有相同方差的分布中， Gaussian分布 $q(x)$ 的熵最大。

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$h(q(x)) = \frac{1}{2}\log(2\pi e\sigma^2)$$

$$D_{\mathrm{KL}}(p||q) = \int \log\left(\frac{p(x)}{q(x)}\right) p(x)\,dx$$

$$= \int p(x)[\log p(x) - \log q(x)]\,dx$$

$$= \int \underbrace{p(x)\log p(x)\,dx}_{-h(p(x))} - \int p(x)\log q(x)\,dx = -h(p(x)) - \int p(x)\log q(x)\,dx$$

# 关于Gaussian分布

还有一个很好的性质， Gaussian分布是最小假设（最大熵假设）

假设： $p(x)$ 是一个均值为 $\mu$，方差为 $\sigma^2$ 的任意概率分布，

证明：在所有具有相同方差的分布中， Gaussian分布 $q(x)$ 的熵最大。

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$D_{\mathrm{KL}}(p||q) = -h(p(x)) - \underbrace{\int p(x) \log q(x)\, dx}$$

$$h(q(x)) = \frac{1}{2}\log(2\pi e\sigma^2)$$

$$= -\frac{1}{2\sigma^2}\int p(x)(x-\mu)^2\, dx - \frac{1}{2}\log(2\pi\sigma^2)$$

$$\int p(x)(x-\mu)^2\, dx = \sigma^2$$

$$= -\frac{1}{2} - \frac{1}{2}\log(2\pi\sigma^2)$$

因为 $p(x)$ 均值为$\mu$，方差为$\sigma^2$

# 关于Gaussian分布

还有一个很好的性质， Gaussian分布是最小假设（最大熵假设）

假设： $p(x)$ 是一个均值为 $\mu$，方差为 $\sigma^2$ 的任意概率分布，

证明：在所有具有相同方差的分布中， Gaussian分布 $q(x)$ 的熵最大。

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$h(q(x)) = \frac{1}{2}\log(2\pi e\sigma^2)$$

$$D_{\mathrm{KL}}(p||q) = -h(p(x)) - \int p(x)\log q(x)\,dx$$

$$= -h\big(p(x)\big) + \frac{1}{2} + \frac{1}{2}\log(2\pi\sigma^2)$$

$$= -h\big(p(x)\big) + h(q(x))$$

因为$D_{\mathrm{KL}}(p||q) \geq 0$，所以 $h(q(x)) \geq h\big(p(x)\big)$ 得证。

# 关于Gaussian分布

结论—— Gaussian分布的熵最大：在所有具有相同方差的分布中， Gaussian分布的熵最大

这个结果说明了为什么自然界中许多随机变量近似服从Gaussian分布，如：

- 热力学第二定律：封闭系统的熵总是趋向最大化（在热平衡状态下，气体分子速度服从高斯分布）
- 深度神经网络初始化：基于Gaussian分布的Xavier或Kaiming初始化，最大熵分布意味引入最少先验
- 等等。。。

# 今日小结

- 贝叶斯法则（后验、似然、先验）
- 最小错误率决策和最小风险决策
- 似然比（N-P决策）
- ROC曲线
- Gaussian分布的点估计
- 中心极限定理、最大熵原理