正态分布下的贝叶斯分类器

翟婷婷

扬州大学
信息工程（人工智能）学院
zhtt@yzu.edu.cn

2023年春

课程目标

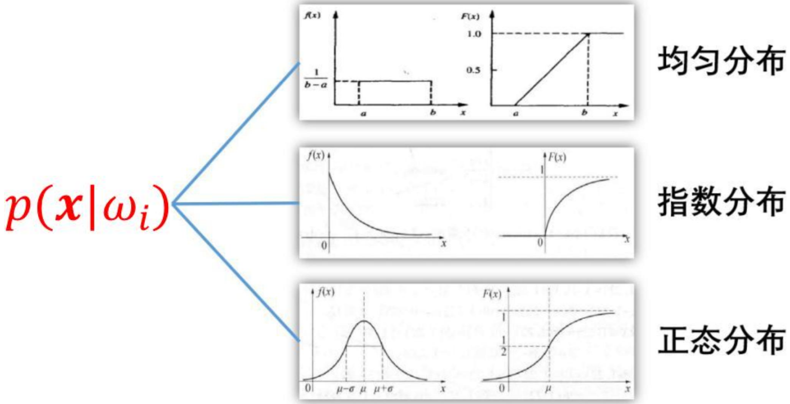
- 了解单变量和多变量正态分布的特点和性质。
- 学会分析正态分布下的最小错误率贝叶斯分类器在不同的分布参数假设下的形式。

回顾贝叶斯分类

- 贝叶斯分类：利用贝叶斯公式，计算待识别样本的特征向量属于每一个类的后验概率，然后将样本分到后验概率最大的那个类中。
- 在计算后验概率时，需要用到每类样本出现的先验概率$P(\omega_i)$和每类样本取到各种特征向量值的类条件概率密度$p(\boldsymbol{x}|\omega_i)$。
- 先验概率可以从大量的重复实验或数据中统计得到，或者，采用均匀先验，即取各类的先验概率相等。
- 类条件概率密度$p(\boldsymbol{x}|\omega_i)$反应每类中特征向量$\boldsymbol{x}$取值的分布情况。经常假设$p(\boldsymbol{x}|\omega_i)$是某种分布的概率密度。
- 当类条件概率密度$p(\boldsymbol{x}|\omega_i)$具有不同的分布形式，那么贝叶斯分类决策规则也就会出现不同的形式。

常见的概率分布形式

- 常见概率分布形式有均匀分布，指数分布和正态分布等。



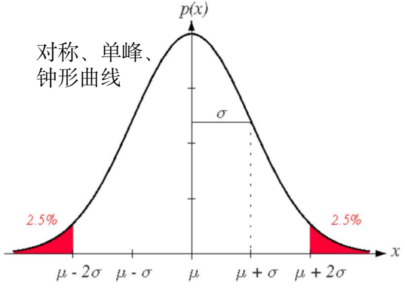
- 正态分布(Normal Distribution)是自然界中最常见的概率分布形式。

正态分布(高斯分布)

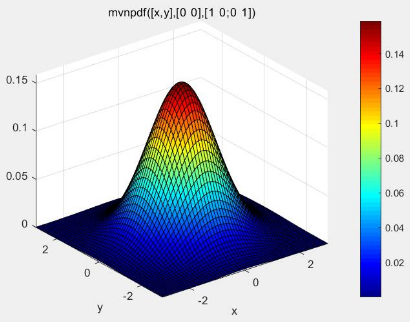
➢ 正态分布：

✓ 单变量正太分布：一维的正态分布，反应一个实值随机变量的取值分布情况。

✓ 多变量正态分布：多维的正态分布，反应一个实值随机向量的取值分布情况。



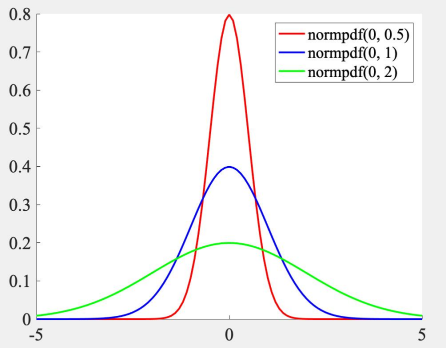
一维正态分布pdf



二维正态分布pdf

单变量正态分布

➢ 单变量正态分布最重要的两个参数是均值和方差。均值反映了分布的中心，方差反映了距离均值的偏离程度。



①在均值μ处的概率密度最大。

② σ决定概率密度曲线的跨度：σ越大，跨度越大；σ越小，跨度越小，函数越尖。

③ σ越小，由高斯分布采样得到的点离均值μ越近。

```
figure; hold on;
plot([-5:0.1:5], normpdf([-5:0.1:5], 0, 0.5),'-r');
plot([-5:0.1:5], normpdf([-5:0.1:5], 0, 1),'-b');
plot([-5:0.1:5], normpdf([-5:0.1:5], 0, 2),'-g');
legend('normpdf(0, 0.5)','normpdf(0, 1)','normpdf(0, 2)');
```

正态分布的概率密度函数

- 单变量正态分布：

  $x \in \mathbb{R}$, $x \sim \mathcal{N}(\mu, \delta^2)$，其概率密度函数为：

$$p(x) = \frac{1}{\delta\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\delta^2}\right]$$

- 多变量正态分布：

  $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$为对称正定矩阵,其概率密度函数为：

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}\,|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

$\boldsymbol{x} = (x^{(1)}, x^{(2)}, \cdots x^{(d)})$，$d$ 维随机向量

$\boldsymbol{\mu} = (\mu^{(1)}, \mu^{(2)}, \cdots \mu^{(d)})$，$d$ 维均值向量

$\boldsymbol{\Sigma}$: $d \times d$的协方差矩阵，$\boldsymbol{\Sigma}_{ij} = \boldsymbol{cov}(x^{(i)}, x^{(j)})$

$|\boldsymbol{\Sigma}|$和$\boldsymbol{\Sigma}^{-1}$为$\boldsymbol{\Sigma}$的行列式和逆矩阵

参数总数：

$$d + \frac{d(d+1)}{2}$$

正态分布的关键参数

- 单变量正态分布：

$$\mu = \mathbb{E}[x] = \int_{-\infty}^{+\infty} xp(x)dx$$

$$\delta^2 = \mathbb{E}[(x-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 p(x)dx$$

- 多变量正态分布：

$$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{x}] = \int \boldsymbol{x}p(\boldsymbol{x})d\boldsymbol{x}$$

$$\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top = \int (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top p(\boldsymbol{x})d\boldsymbol{x}$$

多变量正态分布的数学性质（了解）

- 性质1: 如果一个随机向量$\boldsymbol{x} = (x^{(1)}, x^{(2)}, \cdots x^{(d)})$服从多变量正态分布，则该随机向量的$d$个分量的任意线性组合都符从单变量正态分布，即$Y = a_1 x^{(1)} + a_2 x^{(2)} \cdots + a_d x^{(d)}$服从单变量正态分布，其中$a_1, a_2 \cdots a_d$是任意常数。

- 性质2: 设一个随机向量$\boldsymbol{x} = (x^{(1)}, x^{(2)}, \cdots x^{(d)})$服从多变量正态分布$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则该随机向量的任意两个不相关的分量一定是相互独立的，即若$\boldsymbol{\Sigma}_{ij} = 0$，则随机变量$x^{(i)}$和$x^{(j)}$是相互独立的。

- 问题：多变量高斯分布的等概率密度轨迹是什么样的?
  等概率密度轨迹：概率密度值相等的点构成的轨迹。

多变量正态分布的等概率密度轨迹

- $p(\boldsymbol{x})$的最大值为$p(\boldsymbol{\mu}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}}$
- 记$c \in (0,1]$为一个常数，任何一个概率密度值都可以写成$c \cdot p(\boldsymbol{\mu})$，计算等密度的点满足的等式：

$$p(\boldsymbol{x}) = c \cdot p(\boldsymbol{\mu})$$

- 我们可以得到：

$$(\boldsymbol{x} - \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = -2 \ln c = \ln \frac{1}{c^2}$$

- $\boldsymbol{\Sigma}$是实对称的正定矩阵：$\boldsymbol{\Sigma} = \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{-1}$ (矩阵的特征分解)

  其中$\boldsymbol{Q}$的第i列是$\boldsymbol{\Sigma}$的第i个特征向量，且$\boldsymbol{Q}$是正定阵($\boldsymbol{Q}^{-1} = \boldsymbol{Q}^{\top}$)，$\boldsymbol{\Lambda}$是$\boldsymbol{\Sigma}$的特征值构成的对角矩阵，$\boldsymbol{\Lambda}_{ii} = \lambda_i$。
- 代入：

$$\left(\boldsymbol{Q}^{\top} (\boldsymbol{x} - \boldsymbol{\mu})\right)^{\top} \boldsymbol{\Lambda}^{-1} \boldsymbol{Q}^{\top} (\boldsymbol{x} - \boldsymbol{\mu}) = -2 \ln c = \ln \frac{1}{c^2}$$

多变量正态分布的等概率密度轨迹

- 令$\boldsymbol{y} = \boldsymbol{Q}^{\top}(\boldsymbol{x} - \boldsymbol{\mu})$得到

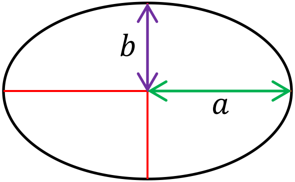
$$\boldsymbol{y}^{\top}\boldsymbol{\Lambda}^{-1}\boldsymbol{y} = \ln\frac{1}{c^2}$$

也就是:

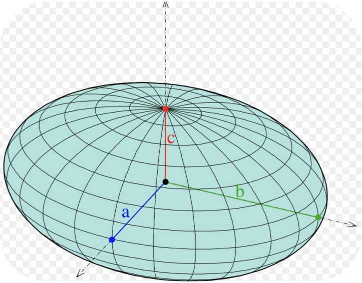
$$\sum_{i=1}^{d}\frac{y_i^2}{\lambda_i} = \ln\frac{1}{c^2}$$

这是超椭球面的方程。

$\boldsymbol{\Sigma}$的特征向量决定主轴的方向， $\boldsymbol{\Sigma}$的特征值决定主轴的长度。

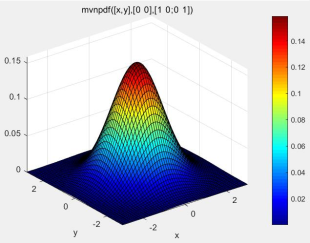


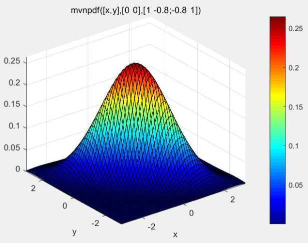
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

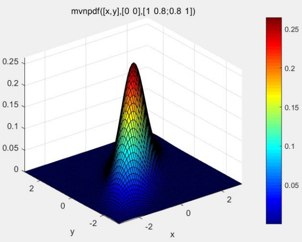


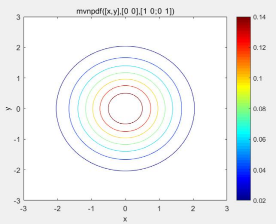
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

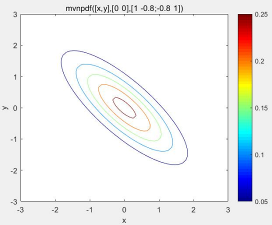
二变量高斯分布的等概率密度轨迹

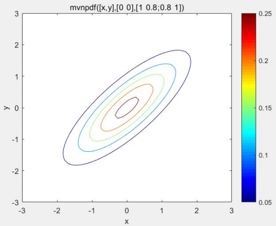













```
figure; ezsurf(inline('mvnpdf([x,y],[0 0],[1 0;0 1])'),[-8 8])
figure;ezcontour(inline('mvnpdf([x,y],[0 0],[1 0;0 1])'),[-3 3]); colormap jet
```

贝叶斯分类规则的等价形式

- 对于一个c类的分类问题，贝叶斯分类规则：

$$\forall j \neq i, P(\omega_i|\boldsymbol{x}) > P(\omega_j|\boldsymbol{x}), \text{则把}\boldsymbol{x}\text{分为}\omega_i\text{类}$$

$$\Updownarrow \ (\Downarrow \quad \Uparrow)$$

$$\forall j \neq i,\ P(\omega_i)p(\boldsymbol{x}|\omega_i) > P(\omega_j)p(\boldsymbol{x}|\omega_j), \text{则把}\boldsymbol{x}\text{分为}\omega_i\text{类}$$

$$\Updownarrow \quad \ln(\cdot)\text{函数是单调递增的}$$

$$\forall j \neq i,\ \ln p(\boldsymbol{x}|\omega_i) + \ln P(\omega_i) > \ln p(\boldsymbol{x}|\omega_j) + \ln P(\omega_j),$$

$$\text{则把}\boldsymbol{x}\text{分为}\omega_i\text{类}$$

- 上述3种分类规则是等价的，都是具有最小错误率的贝叶斯分类规则。

用判别函数表示贝叶斯分类规则

- 对于一个c类的分类问题，定义c个判别函数：

$$g_i(\boldsymbol{x}), i = 1, 2, \cdots c$$

一种常用的模式分类规则就是：

$$\forall j \neq i,\ g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}),\ \text{则把}\boldsymbol{x}\text{分为}\omega_i\text{类。}$$

这种分类形式的分类决策面为：$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$

- 贝叶斯分类规则可以用上述规则表示，取判别函数为：

$$g_i(\boldsymbol{x}) = P(\omega_i|\boldsymbol{x})$$

$$\text{或 } g_i(\boldsymbol{x}) = P(\omega_i)p(\boldsymbol{x}|\omega_i)$$

$$\text{或 } g_i(\boldsymbol{x}) = \ln p(\boldsymbol{x}|\omega_i) + \ln P(\omega_i)$$

这3种判别函数取得相同的分类结果，称它们为最小化错误率的判别函数。

正态分布下的贝叶斯分类

- 在贝叶斯分类决策中，取判别函数：

$$g_i(\boldsymbol{x}) = \ln p(\boldsymbol{x}|\omega_i) + \ln P(\omega_i)\,, i = 1, 2, \cdots c$$

- 如果类条件概率密度$p(\boldsymbol{x}|\omega_i)$是多变量正态分布的概率密度，贝叶斯分类规则会不会更简化呢？
- 设ω_i类中的特征向量取值服从$\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$正态分布：

$$p(\boldsymbol{x}|\omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}}\,|\boldsymbol{\Sigma}_i|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right]$$

代入$g_i(\boldsymbol{x})$可得到：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) \underline{-\frac{d}{2}\ln 2\pi} - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

出现每一类的判别函数中，不影响比较结果

正态分布下的贝叶斯分类：case 1

- 判别函数简化为：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\top}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

- 在不同的分布参数假设下，分析$g_i(\boldsymbol{x})$的取值：

①每类正态分布的**协方差矩阵均相等**，每类中**各个维度的特征是相互独立的且方差相同**，每类样本的先验概率相等，即$\boldsymbol{\Sigma}_i = \delta^2\boldsymbol{I}$，$i = 1,2\cdots c$　　($\boldsymbol{I}$是单位阵)

$P(\omega_i) = 1/c$，$i = 1,2\cdots c$

代入$g_i(\boldsymbol{x})$整理得到：

$$g_i(\boldsymbol{x}) = -\frac{1}{2\delta^2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\top}(\boldsymbol{x}-\boldsymbol{\mu}_i) \underline{-\ln c - d\ln\delta}$$

在所有的判别函数中都有，不影响比较结果

正态分布下的贝叶斯分类：case 1

- 判别函数简化为：

$$g_i(\boldsymbol{x}) = -\frac{1}{2\delta^2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}_i) = -\frac{1}{2\delta^2}||\boldsymbol{x}-\boldsymbol{\mu}_i||^2$$

- 贝叶斯分类规则为：

$$令 i^* = \underset{i}{\mathrm{argmax}}\, g_i(\boldsymbol{x})，\ 把\boldsymbol{x}分为\omega_{i^*}类$$

- 因为 $\underset{i}{\mathrm{argmax}}\, g_i(\boldsymbol{x}) = \underset{i}{\mathrm{argmin}}\, ||\boldsymbol{x}-\boldsymbol{\mu}_i||^2$

- 分类规则本质上变为：

$$令 i^* = \underset{i}{\mathrm{argmin}}\, ||\boldsymbol{x}-\boldsymbol{\mu}_i||^2，\ 把\boldsymbol{x}分为\omega_{i^*}类$$

即$\boldsymbol{x}$离哪一类的均值向量最近，就把$\boldsymbol{x}$分到哪一类中。
这样的分类器称为"最小距离分类器"，(欧式距离)。

正态分布下的贝叶斯分类：case 1

- 现在分析这种情况下的分类决策面是什么样的。
- 第i类和第j类之间的分类决策面方程满足：

$$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$$

将 $g_i(\boldsymbol{x})$和 $g_j(\boldsymbol{x})$代入，并将方程整理成

$$\boldsymbol{w}^{\top}(\boldsymbol{x} - \boldsymbol{b}) = 0$$

形式，得到

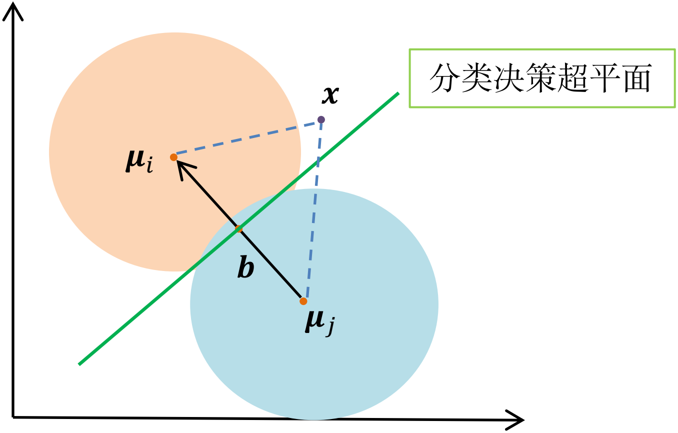
$$\boldsymbol{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$$

$$\boldsymbol{b} = \frac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right)$$

超平面的法向量$\boldsymbol{w}$：从$\boldsymbol{\mu}_j$指向到$\boldsymbol{\mu}_i$

$\boldsymbol{b}$位于$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间连线的中点，分类决策面通过$\boldsymbol{b}$

正态分布下的贝叶斯分类：case 1



分类决策面与$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间的连线垂直，且通过$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间连线的中点。

正态分布下的贝叶斯分类：case 2

- 判别函数为：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

- 在不同的分布参数假设下，分析$g_i(\boldsymbol{x})$的取值：

  ②每类的协方差矩阵相等，每类中各个维度的特征相互独立且方差相同，但各类样本的先验概率不相等，即 $\boldsymbol{\Sigma}_i = \delta^2\boldsymbol{I}$，$i = 1,2\cdots c$　　($\boldsymbol{I}$是单位阵)

带入$g_i(\boldsymbol{x})$整理得到：

$$g_i(\boldsymbol{x}) = \frac{1}{\delta^2}\boldsymbol{\mu}_i{}^\top\boldsymbol{x} - \frac{1}{2\delta^2}\boldsymbol{\mu}_i{}^\top\boldsymbol{\mu}_i - \underline{\frac{1}{2\delta^2}\boldsymbol{x}^\top\boldsymbol{x}} + \ln P(\omega_i) \underline{- d\ln\delta}$$

在所有的判别函数中都有，不影响比较结果

正态分布下的贝叶斯分类：case 2

- 判别函数简化为：

$$g_i(\boldsymbol{x}) = \frac{1}{\delta^2}\boldsymbol{\mu}_i{}^{\top}\boldsymbol{x} - \frac{1}{2\delta^2}\boldsymbol{\mu}_i{}^{\top}\boldsymbol{\mu}_i + \ln P(\omega_i)$$

看到$g_i(\boldsymbol{x})$是线性判别函数。

- 第i类和第j类间的分类决策面为：$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$

将$g_i(\boldsymbol{x})$和$g_j(\boldsymbol{x})$代入，注意$\frac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^{\top}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{||\boldsymbol{\mu}_i - \boldsymbol{\mu}_j||^2} = 1$

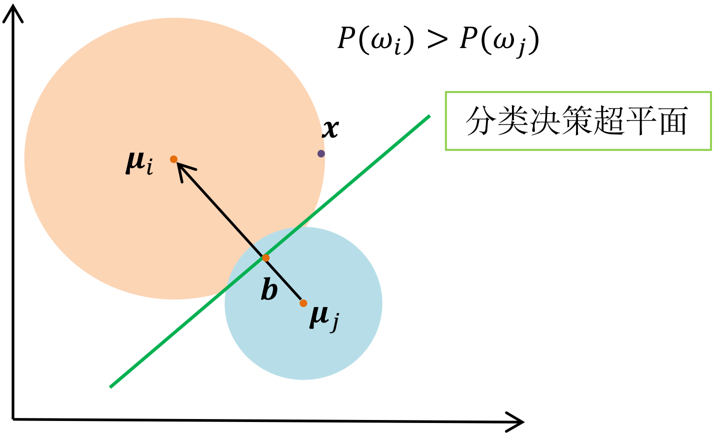
可将方程整理成 $\boldsymbol{w}^{\top}(\boldsymbol{x} - \boldsymbol{b}) = 0$，其中

$$\boldsymbol{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$$

$$\boldsymbol{b} = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\delta^2 \ln \frac{P(\omega_i)}{P(\omega_j)}}{||\boldsymbol{\mu}_i - \boldsymbol{\mu}_j||^2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

$P(\omega_i) > P(\omega_j) \Rightarrow \ln\frac{P(\omega_i)}{P(\omega_j)} > 0 \Rightarrow \boldsymbol{b}$距离$\boldsymbol{\mu}_j$更近一点

正态分布下的贝叶斯分类：case 2



分类决策面与$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间的连线垂直，但会离开先验概率大的那一类，使得更多样本被识别为先验概率大的那一类。

正态分布下的贝叶斯分类：case 3

- 判别函数为：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

- 在不同的分布参数假设下，分析$g_i(\boldsymbol{x})$的取值：

  ③每类的协方差矩阵相等，但每类中各个维度的特征不相互独立，各类样本的先验概率相等，即

  $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$, $i = 1,2\cdots c$　　($\boldsymbol{I}$是单位阵)

  $P(\omega_i) = 1/c$, $i = 1,2\cdots c$

  带入$g_i(\boldsymbol{x})$得到：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) \underline{- \frac{1}{2}\ln|\boldsymbol{\Sigma}| - \ln c}$$

在所有的判别函数中都有，不影响比较结果

正态分布下的贝叶斯分类：case 3

- 判别函数简化为：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)$$

- 贝叶斯分类规则本质上变为：

令$i^* = \arg\min_i (\boldsymbol{x}-\boldsymbol{\mu}_i)^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)$，把$\boldsymbol{x}$分为$\omega_{i^*}$类

这样的分类器是“马氏最小距离分类器”，注意：距离指的是马氏距离，而不是欧式距离。

马氏距离是由印度统计学家马哈拉诺比斯(P. C. Mahalanobis)提出的一种距离度量，它与欧式距离不同的是它考虑到各维度特征之间的相关性，并且是尺度无关的(scale-invariant)，即独立于测量尺度。

正态分布下的贝叶斯分类：case 3

- 第i类和第j类间的分类决策面为$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$
  将$g_i(\boldsymbol{x})$和$g_j(\boldsymbol{x})$代入并将方程整理成$\boldsymbol{w}^\top(\boldsymbol{x} - \boldsymbol{b}) = 0$
  得到
  $$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$
  $$\boldsymbol{b} = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$
  $\boldsymbol{b}$位于$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间连线的中点，分类决策面通过$\boldsymbol{b}$。

- 分类决策面不再与$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间的连线垂直，但是会通过$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间连线的中点。

正态分布下的贝叶斯分类：case 4

- 判别函数为：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

- 在不同的分布参数假设下，分析$g_i(\boldsymbol{x})$的取值：

④每类的协方差矩阵相等，但每类中各个维度的特征不相互独立，各类样本的先验概率不相等，即

$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$，$i = 1,2\cdots c$ （$\boldsymbol{I}$是单位阵）

带入$g_i(\boldsymbol{x})$得到：

$$g_i(\boldsymbol{x}) = \boldsymbol{\mu}_i{}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i \underbrace{- \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}\ln|\boldsymbol{\Sigma}|}_{} + \ln P(\omega_i)$$

在所有的判别函数中都有，不影响分类结果

正态分布下的贝叶斯分类：case 4

- 判别函数简化为：

$$g_i(\boldsymbol{x}) = \boldsymbol{\mu}_i{}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln P(\omega_i)$$

看到$g_i(\boldsymbol{x})$仍是线性判别函数。

- 第i类和第j类间的分类决策面为：$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$

将 $g_i(\boldsymbol{x})$和 $g_j(\boldsymbol{x})$代入整理成 $\boldsymbol{w}^{\top}(\boldsymbol{x} - \boldsymbol{b}) = 0$，其中

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

$$\boldsymbol{b} = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\boldsymbol{\mu}_i - \boldsymbol{\mu}_j}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)} \ln \frac{P(\omega_i)}{P(\omega_j)}$$

- 分类决策超平面会偏向先验概率较小的那一类。

正态分布下的贝叶斯分类：case 5

- 判别函数为：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

- 在不同的分布参数假设下，分析$g_i(\boldsymbol{x})$的取值：

  ⑤每类的协方差矩阵不相等，每类中各个维度的特征不相互独立，各类样本的先验概率不相等。

对$g_i(\boldsymbol{x})$整理得到：

$$g_i(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{x} + \boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

正态分布下的贝叶斯分类：case 5

- 判别函数可以变形为：

$$g_i(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{W}_i \boldsymbol{x} + \boldsymbol{w}_i^\top \boldsymbol{x} + b_i$$

其中，

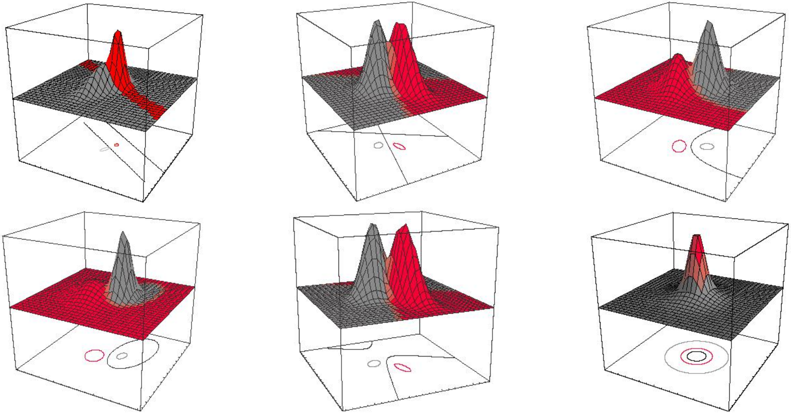
$$\boldsymbol{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1}$$

$$\boldsymbol{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

$$b_i = -\frac{1}{2}\boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

- 此时，分类决策面是超二次曲面，会随着$\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i$的不同而呈现为不同形式。贝叶斯分类器在这种情况下是非线性分类器。

正态分布下的贝叶斯分类：case 5



分类决策超曲面投影到2维平面上：双直线，交叉直线，抛物线，椭圆，双曲线，椭圆。

小结

- 正态分布，也称高斯分布，是自然界中最常见的分布形式。多变量正态分布是将一维的正态分布扩展到高维的情况中，反应一个实值随机向量的取值分布情况。
- 多变量正态分布：$\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，概率密度函数为：

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

- 多变量正态分布中各个维度的任意线性组合都服从单变量正态分布。
- 多变量正态分布中，任意两个维度之间的不相关性等价于相互独立性。
- 多变量正态分布的等概率密度轨迹是一个超椭球体，在二维情况下，是一个椭圆。

小结

- 一种常见的表示模式分类器的方式是：对于一个c类的分类问题，定义c个判别函数：

$$g_i(\boldsymbol{x}), i = 1, 2, \cdots c$$

采用分类规则：

$$\forall j \neq i,\ g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}),\ 则把\boldsymbol{x}分为\omega_i类。$$

这种分类形式的分类决策面为：$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$

- 最小错误率的贝叶斯分类器等价于在上述模式分类方式中采用判别函数：

$$g_i(\boldsymbol{x}) = \ln p(\boldsymbol{x}|\omega_i) + \ln P(\omega_i), i = 1, 2, \cdots c$$

重点

- 当类条件概率密度$p(\boldsymbol{x}|\omega_i)$为多变量正态分布的概率密度函数时，贝叶斯分类规则的形式可以进一步简化。

小结

- 在贝叶斯分类中，设ω_i类样本的特征向量取值服从多变量的正态分布$\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$：

难点

  - ✓ 当$\boldsymbol{\Sigma}_i = \delta^2 \boldsymbol{I}$，$P(\omega_i) = 1/c$，$i = 1,2 \cdots c$

    贝叶斯分类器是一个线性分类器，且第i类与第j类之间的分类决策超平面垂直平均$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间的连线。

  - ✓ 当$\boldsymbol{\Sigma}_i = \delta^2 \boldsymbol{I}$，$i = 1,2 \cdots c$

    贝叶斯分类器是一个线性分类器，且第i类与第j类之间的分类决策超平面与$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间的连线垂直，偏向先验概率小的那一类。

小结

- ✓ 当$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$，$P(\omega_i) = 1/c$，$i = 1,2\cdots c$

  贝叶斯分类器仍然是一个线性分类器，且第i类与第j类之间的分类决策超平面不再与$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间的连线垂直，但通过$\boldsymbol{\mu}_i$与$\boldsymbol{\mu}_j$之间连线的中点。

- ✓ 当$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$，$i = 1,2\cdots c$

  贝叶斯分类器仍然是一个线性分类器，且第i类与第j类之间的分类决策超平面不再与$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$之间的连线垂直，偏向先验概率小的那一类。

- ✓ 当$\boldsymbol{\Sigma}_i =$任意，$i = 1,2\cdots c$

  贝叶斯分类器是一个非线性分类器，分类决策面是超二次曲面。