一、第二章作业

■ 设以下模式类别具有正态概率密度函数:

ω_1: $\{(0\ 0)^T, (2\ 0)^T, (2\ 2)^T, (0\ 2)^T\}$

ω_2: $\{(4\ 4)^T, (6\ 4)^T, (6\ 6)^T, (4\ 6)^T\}$

(1) 设 $P(\omega_1)= P(\omega_2)=1/2$, 求这两类模式之间的贝叶斯判别界面的方程式。

(2) 绘出判别界面。

$$f_{\mathbf{x}}(x_1,\ldots,x_k)=\frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

2.1 贝叶斯判别原则

■贝叶斯判别　$P(\omega_1 \mid \boldsymbol{x})$　?　$P(\omega_2 \mid \boldsymbol{x})$，

$$P(\omega_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\omega_i)P(\omega_i)}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\omega_i)P(\omega_i)}{\sum_{i=2}^{2} p(\boldsymbol{x}|\omega_i)P(\omega_i)}$$

$$p(\boldsymbol{x}|\omega_1)P(\omega_1) > p(\boldsymbol{x}|\omega_2)P(\omega_2) \quad \Rightarrow \quad \boldsymbol{x} \in \omega_1$$

$$p(\boldsymbol{x}|\omega_2)P(\omega_2) > p(\boldsymbol{x}|\omega_1)P(\omega_1) \quad \Rightarrow \quad \boldsymbol{x} \in \omega_2$$

$$d_1(\boldsymbol{x}) = lnP(\omega_1) - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1)$$

$$d_2(\boldsymbol{x}) = lnP(\omega_2) - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_2)$$

$$m_1 = \frac{1}{4}\sum_{j=1}^{4} x_{1j} = \frac{1}{4}[(0,0)^T + (2,0)^T + (2,2)^T + (0,2)^T] = (1,1)^T$$

$$m_2 = \frac{1}{4}\sum_{j=1}^{4} x_{2j} = \frac{1}{4}[(4,4)^T + (6,4)^T + (6,6)^T + (4,6)^T] = (5,5)^T$$

$$C_1 = \frac{1}{4}\sum_{j=1}^{4}(x_{1j} - m_1)(x_{1j} - m_1)^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C_2 = \frac{1}{4}\sum_{j=1}^{4}(x_{2j} - m_2)(x_{2j} - m_2)^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\therefore\ C_1 = C_2 = C = I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad C^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

判别界面为:

$$d_1(x) - d_2(x)$$

$$= lnP(w_1) - lnP(w_2) + (m_1 - m_2)^T C^{-1} x - \frac{1}{2} m_1^T C^{-1} m_1 + \frac{1}{2} m_2^T C^{-1} m_2$$

$$= 0$$

代入可得:

$$x_1 + x_2 = 6$$

1. 假设国庆节假期你到怀柔山区去徒步。你注意到道路两旁到处都是野生浆果。我们都知道有些浆果有毒，有些可以食用。下表给出了一些达人传授的知识（注：该表纯属虚构，不反映现实生活）。

样本索引	是否是黑红色 **R**	是否成簇生长 **C**	是否有斑点 **S**	是否有毒 **Y**
1	0	0	1	1
2	1	1	0	1
3	0	1	1	1
4	0	0	0	0
5	0	1	0	0
6	1	0	1	0
7	0	0	1	1
8	1	0	0	1

(1) 根据上述数据（1-8）训练朴素贝叶斯分类器。计算给定类别的情况下各属性取值的条件概率，例如$P(\mathrm{R}=1|Y=1)$、$P(\mathrm{R}=1|Y=0)$，以及各类的先验概率$P(Y=1)$、$P(Y=0)$。

解：类别先验：$P(Y=0)=\frac{3}{8}$、$P(Y=1)=\frac{5}{8}$

各属性值的类条件概率估计采用 Laplace 平滑估计

$P(R=0|Y=0)=\frac{2+1}{3+2}=\frac{3}{5}$、$P(R=1|Y=0)=\frac{1+1}{3+2}=\frac{2}{5}=1-P(R=0|Y=0)$

$P(R=0|Y=1)=\frac{3+1}{5+2}=\frac{4}{7}$、$P(R=1|Y=1)=\frac{3}{7}$

$P(C=0|Y=0)=\frac{2+1}{3+2}=\frac{3}{5}$、$P(C=1|Y=0)=\frac{2}{5}$

$P(C=0|Y=1)=\frac{3+1}{5+2}=\frac{4}{7}$、$P(C=1|Y=1)=\frac{3}{7}$

$P(S=0|Y=0)=\frac{2+1}{3+2}=\frac{3}{5}$、$P(S=1|Y=0)=\frac{2}{5}$

$P(S=0|Y=1)=\frac{2+1}{5+2}=\frac{3}{7}$、$P(S=1|Y=1)=\frac{4}{7}$

(2) 给定如下 1 个样本，根据（1）中训练好的朴素贝叶斯分类器，预测它们是否有毒。

样本索引	是否是黑红色 **R**	是否成簇生长 **C**	是否有斑点 **S**	是否有毒 **Y**
9	1	0	0	？

答：

$$P(Y=0|R=1,C=0,S=0)\propto P(Y=0)P(R=1|Y=0)P(C=0|Y=0)\,P(S=0|Y=0)$$

$$=\frac{3}{8}\times\frac{2}{5}\times\frac{3}{5}\times\frac{3}{5}=\frac{27}{500}\approx 0.054$$

$$P(Y=1|R=1,C=0,S=0)\propto P(Y=1)P(R=1|Y=1)P(C=0|Y=1)\,P(S=0|Y=1)$$

$$=\frac{5}{8}\times\frac{3}{7}\times\frac{4}{7}\times\frac{3}{7}=\frac{45}{686}\approx 0.066$$

$$P(Y=0|R=1,C=0,S=0)=\frac{0.054}{0.054+0.066}=0.45$$

$$P(Y=1|R=1,C=0,S=0)=1-P(Y=0|R=1,C=0,S=0)=0.55$$

$$P(Y=0|R=1,C=0,S=0)>P(Y=1|R=1,C=0,S=0)$$

所以该浆果有毒。

（3）假设将有毒的浆果误判为可食用的代价是 20，将可食用的浆果误判为有毒的代价是 10，请根据最小风险原则，判断上述两种浆果是否可食用。

解：根据题意，

$L_{00} = 0$，$L_{01} = 10$

$L_{10} = 20$，$L_{11} = 0$

$R(\hat{y} = 0|\boldsymbol{x}) = L_{00}P(Y = 0|\boldsymbol{x}) + L_{10}P(Y = 1|\boldsymbol{x}) = 0.45\times0 + 0.55\times20 = 1.1$

$R(\hat{y} = 1|\boldsymbol{x}) = L_{01}P(Y = 0|\boldsymbol{x}) + L_{11}P(Y = 1|\boldsymbol{x}) = 0.45\times10 + 0.55\times0 = 4.5$

$R(\hat{y} = 0|\boldsymbol{x}) > R(\hat{y} = 1|\boldsymbol{x})$

所以判定该浆果有毒。

$$r_j = \sum_{i=1}^{M} L_{ij}P(\omega_i|x)$$

- L_{ij}称为将本应属于ω_i类的样本判别成属于ω_j类的是非代价。

二、第三章作业

一、在一个 10 类的模式识别问题中，有 3 类单独满足多类情况 1，其余的类别满足多类情况 2。问该模式识别问题所需判别函数的最少数目是多少？

解答：

多类情况 1：把 M 类多类问题分成 M 个两类问题，因此共有 M 个判别函数；

多类情况 2：要分开 M 类模式，共需 M(M-1)/2 个判别函数。

所以，该模式识别问题所需判别函数的最少数目是

$3 + 7 * (7-1)/2 = 24$

二、第三章作业

二、 一个三类问题，其判别函数如下：

$d_1(\mathbf{x})=-x_1,\ d_2(\mathbf{x})=x1+x_2-1,\ d_3(\mathbf{x})=x_1-x_2-1$

1.设这些函数是在多类情况 1 条件下确定的，绘出其判别界面和每一个模式类别的区域。

2.设为多类情况 2，并使：$d_{12}(\mathbf{x})= d_1(\mathbf{x}),\ d_{13}(\mathbf{x})= d_2(\mathbf{x}),\ d_{23}(\mathbf{x})= d_3(\mathbf{x})$。绘出其判别界面和多类情况 2 的区域。

设 $d_1(\mathbf{x}), d_2(\mathbf{x})$和 $d_3(\mathbf{x})$是在多类情况 3 的条件下确定的，绘出其判别界面和每类的区域。

略

二、第三章作业

三、两类模式，每类包括 5 个 3 维不同的模式向量，且良好分布。如果它们是线性可分的，问权向量至少需要几个系数分量？假如要建立二次的多项式判别函数，又至少需要几个系数分量？（设模式的良好分布不因模式变化而改变。）

对于*n*维*x*向量，若用*r*次多项式，*d*(*x*)的权系数的总项数为：

$$N_w = C_{n+r}^{r} = \frac{(n+r)!}{r!\,n!}$$

解答:

如果它们是线性可分的，权向量至少需要 4 个系数分量；

假如要建立二次的多项式判别函数，至少需要 C_{3+2}^{2} = 10个系数分量。

二、第三章作业

四、

1．用感知器算法求下列模式分类的解向量 $\boldsymbol{w}$:

ω_1: $\{(0\ 0\ 0)^T, (1\ 0\ 0)^T, (1\ 0\ 1)^T, (1\ 1\ 0)^T\}$

ω_2: $\{(0\ 0\ 1)^T, (0\ 1\ 1)^T, (0\ 1\ 0)^T, (1\ 1\ 1)^T\}$

感知器算法可统一写成： $w(k+1) = \begin{cases} w(k) & \text{, if } w^T(k)x^k > 0 \\ w(k) + Cx^k & \text{if } w^T(k)x^k \le 0 \end{cases}$

将属于ω_2的训练样本乘以（-1），并写成增广向量的形式。

$x_{①}=(0\ \ 0\ \ 0\ \ 1)^T$，$x_{②}=(1\ \ 0\ \ 0\ \ 1)^T$，$x_{③}=(1\ \ 0\ \ 1\ \ 1)^T$，$x_{④}=(1\ \ 1\ \ 0\ \ 1)^T$

$x_{⑤}=(0\ \ 0\ \ -1\ \ -1)^T$，$x_{⑥}=(0\ \ -1\ \ -1\ \ -1)^T$，$x_{⑦}=(0\ \ -1\ \ 0\ \ -1)^T$，$x_{⑧}=(-1\ \ -1\ \ -1\ \ -1)^T$

第一轮迭代：取 C=1，$\boldsymbol{w}(1)=(0\ \ 0\ \ 0\ \ 0)^T$

因 $\boldsymbol{w}^T(1)\boldsymbol{x}_{①}=(0\ \ 0\ \ 0\ \ 0)(0\ \ 0\ \ 0\ \ 1)^T=0 \not> 0$，故 $\boldsymbol{w}(2)=\boldsymbol{w}(1)+\boldsymbol{x}_{①}=(0\ \ 0\ \ 0\ \ 1)^T$

因 $\boldsymbol{w}^T(2)\boldsymbol{x}_{②}=(0\ \ 0\ \ 0\ \ 1)(1\ \ 0\ \ 0\ \ 1)^T=1>0$，故 $\boldsymbol{w}(3)=\boldsymbol{w}(2)=(0\ \ 0\ \ 0\ \ 1)^T$

因 $\boldsymbol{w}^T(3)\boldsymbol{x}_{③}=(0\ \ 0\ \ 0\ \ 1)(1\ \ 0\ \ 1\ \ 1)^T=1>0$，故 $\boldsymbol{w}(4)=\boldsymbol{w}(3)=(0\ \ 0\ \ 0\ \ 1)^T$

因 $\boldsymbol{w}^T(4)\boldsymbol{x}_{④}=(0\ \ 0\ \ 0\ \ 1)(1\ \ 1\ \ 0\ \ 1)^T=1>0$，故 $\boldsymbol{w}(5)=\boldsymbol{w}(4)=(0\ \ 0\ \ 0\ \ 1)^T$

因 $\boldsymbol{w}^T(5)\boldsymbol{x}_{⑤}=(0\ \ 0\ \ 0\ \ 1)(0\ \ 0\ \ -1\ \ -1)^T=-1 \not> 0$，故 $\boldsymbol{w}(6)=\boldsymbol{w}(5)+\boldsymbol{x}_{⑤}=(0\ \ 0\ \ -1\ \ 0)^T$

因 $\boldsymbol{w}^T(6)\boldsymbol{x}_{⑥}=(0\ \ 0\ \ -1\ \ 0)(0\ \ -1\ \ -1\ \ -1)^T=1>0$，故 $\boldsymbol{w}(7)=\boldsymbol{w}(6)=(0\ \ 0\ \ -1\ \ 0)^T$

因 $\boldsymbol{w}^T(7)\boldsymbol{x}_{⑦}=(0\ \ 0\ \ -1\ \ 0)(0\ \ -1\ \ 0\ \ -1)^T=0 \not> 0$，故 $\boldsymbol{w}(8)=\boldsymbol{w}(7)+\boldsymbol{x}_{⑦}=(0\ \ -1\ \ -1\ \ -1)^T$

因 $\boldsymbol{w}^T(8)\boldsymbol{x}_{⑧}=(0\ \ -1\ \ -1\ \ -1)(-1\ \ -1\ \ -1\ \ -1)^T=3>0$，故 $\boldsymbol{w}(9)=\boldsymbol{w}(8)=(0\ \ -1\ \ -1\ \ -1)^T$

这里，第 1、5、7 步为错误分类，应“罚”。

因为只有对全部模式都能正确判别的权向量才是正确的解，因此需进行第二轮迭代。

二、第三章作业

五、用多类感知器算法求下列模式的判别函数：

ω_1: $(-1\ -1)^T$，　ω_2: $(0\ 0)^T$，　ω_3: $(1\ 1)^T$

第k次迭代时，一个属于ω_i类的样本x送入分类器，则应先计算出M个判别函数：$d_j(k) = w_j(k)x, j = 1, 2, \cdots, M$

若$d_i(k) > d_j(k)$，$j = 1, 2, \ldots, M, \forall j \neq i$的条件成立，则权向量不变，即

$$w_j(k+1) = w_j(k), j = 1, 2, \cdots, M$$

若其中第l个权向量使得$d_i(k) \leq d_l(k)$，则相应的权向量应做调整，即

$$w_i(k+1) = w_i(k) + Cx$$

$$w_l(k+1) = w_l(k) - Cx$$

$$w_j(k+1) = w_j(k), j = 1, 2, \cdots, M\ , j \neq i, j \neq l$$

C是一个正常数。权向量的初始值$w_i(l)$，$i = 1,2,\ldots,M$可视情况任意选择。

将模式样本写成增广形式：

$x_{①}=(-1\ \ -1\ \ 1)^T$， $x_{②}=(0\ \ 0\ \ 1)^T$， $x_{③}=(1\ \ 1\ \ 1)^T$

取初始值$w_1(1)=w_2(1)=w_3(1)=(0\ \ 0\ \ 0)^T$，C=1。

第一轮迭代（k=1）：以 $x_{①}=(-1\ \ -1\ \ 1)^T$作为训练样本

$d_1(1)=w_1^T(1)x_{①}=(0\ \ 0\ \ 0)(-1\ \ -1\ \ 1)^T=0$

$d_2(1)=w_2^T(1)x_{①}=(0\ \ 0\ \ 0)(-1\ \ -1\ \ 1)^T=0$

$d_3(1)=w_3^T(1)x_{①}=(0\ \ 0\ \ 0)(-1\ \ -1\ \ 1)^T=0$

因 $d_1(1)\ngtr d_2(1)$， $d_1(1)\ngtr d_3(1)$， 故

$\boxed{w_1(2)=w_1(1)+x_{①}=(-1\ \ -1\ \ 1)^T}$

$w_2(2)=w_2(1)-x_{①}=(1\ \ 1\ \ -1)^T$

$w_3(2)=w_3(1)-x_{①}=(1\ \ 1\ \ -1)^T$

三、第四章作业

1、 设有如下三类模式样本集 ω_1，ω_2 和 ω_3，其先验概率相等，求 S_w 和 S_b

ω_1: $\{(1\ 0)^T, (2\ 0)^T, (1\ 1)^T\}$

ω_2: $\{(-1\ 0)^T, (0\ 1)^T, (-1\ 1)^T\}$

ω_3: $\{(-1\ -1)^T, (0\ -1)^T, (0\ -2)^T\}$

略

2、设有如下两类样本集，其出现的概率相等：

ω_1: $\{(0\ 0\ 0)^T, (1\ 0\ 0)^T,$
$(1\ 0\ 1)^T, (1\ 1\ 0)^T\}$

ω_2: $\{(0\ 0\ 1)^T, (0\ 1\ 0)^T,$
$(0\ 1\ 1)^T, (1\ 1\ 1)^T\}$

用 K-L 变换，分别把特征空间维数降到二维和一维，并画出样本在该空间中的位置

KL最佳变换-步骤

给定N个样本 $x^1, x^2, \ldots, x^N$ ，利用KL变换，将其降至m维

1.计算样本均值：$\boldsymbol{m}=\mathrm{E}(\boldsymbol{x})$

2.平移样本： $\boldsymbol{z}=\boldsymbol{x}-\boldsymbol{m}$

3.计算z的自相关矩阵$\mathrm{R}(z)=\mathrm{E}(z\, z^{\mathrm{T}})=\mathrm{E}((\boldsymbol{x}-\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}})=C_{\mathrm{x}}$

4.求$\mathrm{R}(z)$（即$\boldsymbol{x}$的协方差矩阵)的特征向量，取最大的m个特征值 对应的m个特征向量构成变换矩阵 Φ

样本变换后的为： $\Phi^T z$

对于测试样本x, 其变换后为： $\boxed{\Phi^T(\boldsymbol{x}-\boldsymbol{m})}$

对x重构：$\tilde{\boldsymbol{x}} = \boldsymbol{\Phi}\boldsymbol{\Phi}^T(\boldsymbol{x}-\boldsymbol{m})+\boldsymbol{m}$

总体均值向量：

第一类的第j个样本

$$m=E\{x\}=0.5*\frac{1}{4}\sum_{j=1}^{4}\boxed{x_{1j}}+0.5*\frac{1}{4}\sum_{j=1}^{4}x_{2j}=\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\end{pmatrix}^T$$

将所有这些样本的各分量都减去 0.5，便可以将所有这些样本的均值移到原点：

$$\omega_1:\ \left\{\begin{pmatrix}-\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T\right\}$$

$$\omega_2:\ \left\{\begin{pmatrix}-\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}-\frac{1}{2} & \frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}-\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\end{pmatrix}^T\right\}$$

此时，符合 K-L 变换进行特征压缩的最佳条件。

因 $P(\omega_1)=P(\omega_2)=0.5$，故

$$R=\sum_{i=1}^{2}P(\omega_i)E\{xx^T\}=\frac{1}{2}\left[\frac{1}{4}\sum_{j=1}^{4}x_{1j}x_{1j}^T\right]+\frac{1}{2}\left[\frac{1}{4}\sum_{j=1}^{4}x_{2j}x_{2j}^T\right]=\begin{pmatrix}\frac{1}{4} & 0 & 0\\ 0 & \frac{1}{4} & 0\\ 0 & 0 & \frac{1}{4}\end{pmatrix}$$

解特征值方程|R−λI|=0，求 R 的特征值，得特征值 $\lambda_1=\lambda_2=\lambda_3=0.25$

其对应的特征向量可由 $R\Phi_i=\lambda_i\Phi_i$ 求得：

$$\Phi_1=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad \Phi_2=\begin{pmatrix}0\\1\\0\end{pmatrix},\quad \Phi_3=\begin{pmatrix}0\\0\\1\end{pmatrix}$$

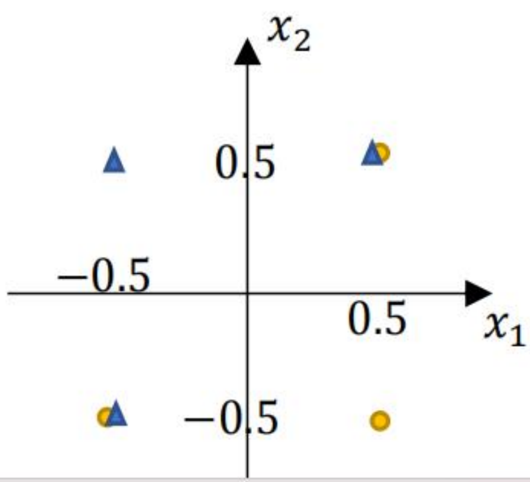
1）　把特征空间维数降到二维：

选 λ_1 和 λ_2 对应的特征向量 Φ_1、Φ_2 作变换矩阵，得到 $\Phi=\begin{pmatrix}1&0\\0&1\\0&0\end{pmatrix}$

由 $y=\Phi^T x$ 得变换后的二维模式特征为：

ω_1: $\left\{\begin{pmatrix}-\frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\end{pmatrix}^T\right\}$

ω_2: $\left\{\begin{pmatrix}-\frac{1}{2} & -\frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}-\frac{1}{2} & \frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}-\frac{1}{2} & \frac{1}{2}\end{pmatrix}^T,\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\end{pmatrix}^T\right\}$



四、第六章作业

生成式分类器和判别式分类器：高斯贝叶斯和逻辑回归

1. 考虑一种特定类别的高斯朴素贝叶斯分类器，其中：

- y是一个布尔变量服从伯努利分布，参数为$\pi = P(y=1)$，因此$P(Y=0) = 1-\pi$.
- $x = [x_1, \ldots, x_D]^T$，其中每个特征 x_i 是一个连续随机变量。对于每个x_i，$P(x_i|y=k)$ 是一个高斯分布 $N(\mu_{ik}, \sigma_i)$。请注意，σ_i是高斯分布的标准差，不依赖于 k。
- 对于所有$i \neq j$，给定y，x_i 和x_j 是条件独立的（即所谓的"朴素"分类器）。

问：上述这种高斯朴素贝叶斯判别器与逻辑回归得到分类器形式是一致的。

逻辑回归及求解

$$P(y=1|\boldsymbol{x}) = f(\boldsymbol{x},\boldsymbol{w}) = g(\boldsymbol{w}^T\boldsymbol{x}) = \frac{1}{1+\exp(-\boldsymbol{w}^T\boldsymbol{x})}$$

- 概率分布： $\boxed{P(y|\boldsymbol{x},\boldsymbol{w}) = (f(\boldsymbol{x},\boldsymbol{w}))^{y}(1-f(\boldsymbol{x},\boldsymbol{w}))^{1-y}}$ *Bernoulli*
- 似然:

$$\boxed{L(\boldsymbol{w}) = \prod_{i=1}^{N} P(y^i|\boldsymbol{x}^i,\boldsymbol{w}) = \prod_{i=1}^{N}\left(f(\boldsymbol{x}^i,\boldsymbol{w})\right)^{y^i}\left(1-f(\boldsymbol{x}^i,\boldsymbol{w})\right)^{1-y^i}}$$

- 最大化log 似然:

$$\boxed{l(\boldsymbol{w}) = \log L(\boldsymbol{w}) = \sum_{i=1}^{N}\left(y^i\log f(\boldsymbol{x}^i,\boldsymbol{w}) + (1-y^i)\log\left(1-f(\boldsymbol{x}^i,\boldsymbol{w})\right)\right)}$$

+正则项

- 梯度:

$$\frac{\partial l(\boldsymbol{w})}{\partial w_j} = (y^i - f(\boldsymbol{x}^i,\boldsymbol{w}))x_j^i, \forall(\boldsymbol{x}^i, y^i)$$

- SGD:

$$w_j = w_j + \alpha(y^i - f(\boldsymbol{x}^i,\boldsymbol{w}))x_j^i$$

GDA

$$P(Y=1|\boldsymbol{x}) = \frac{\boxed{p(\boldsymbol{x}|Y=1)}P(Y=1)}{p(\boldsymbol{x})}$$

$$= \frac{p(\boldsymbol{x}|Y=1)P(Y=1)}{p(\boldsymbol{x}|Y=1)P(Y=1) + p(\boldsymbol{x}|Y=0)P(Y=0)}$$

根据上述特定高斯朴素贝叶斯分类器的假设，以及贝叶斯法则，有：

$$P(Y=1|X)=\frac{P(Y=1)P(X|Y=1)}{P(Y=1)P(X|Y=1)+P(Y=0)P(X|Y=0)}$$

$$=\frac{1}{1+\frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}}=\frac{1}{1+\exp\left(\ln\frac{P(Y=0)P(X|Y=0)}{P(Y=1)P(X|Y=1)}\right)}$$

由给定 Y，x的条件独立性假设，可得：

$$P(Y=1|X)=\frac{1}{1+\exp\left(\ln\frac{P(Y=0)}{P(Y=1)}+\ln\left(\prod_i\frac{P(x_i|Y=0)}{P(x_i|Y=1)}\right)\right)}$$

$$=\frac{1}{1+\exp\left(\ln\frac{1-\pi}{\pi}+\sum_i\ln\frac{P(x_i|Y=0)}{P(x_i|Y=1)}\right)}$$

再根据$P(x_i|Y=y_k)$服从高斯分布$\mathcal{N}(\mu_{ik},\sigma_i)$，可得：

$$\sum_i\ln\frac{P(x_i|Y=0)}{P(x_i|Y=1)}=\sum_i\ln\frac{\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{-(x_i-\mu_{i0})^2}{2\sigma_i^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{-(x_i-\mu_{i1})^2}{2\sigma_i^2}\right)}$$

$$=\sum_i\ln\exp\left(\frac{(x_i-\mu_{i1})^2-(x_i-\mu_{i0})^2}{2\sigma_i^2}\right)\quad=\sum_i\left(\frac{\mu_{i0}-\mu_{i1}}{\sigma_i^2}x_i+\frac{\mu_{i1}^2-\mu_{i0}^2}{2\sigma_i^2}\right)$$

2. 一般高斯朴素分类器和逻辑回归

- 去掉 $P(x_i|y=k)$的标准差σ_i不依赖于k的假设，即对于每个x_i，$P(x_i|y=k)$是一个高斯分布$N(\mu_{ik}, \sigma_{ik})$，其中$i=1, \ldots, D, k=0,1$。

问：这个更一般的高斯朴素贝叶斯分类器所隐含的$P(y|x)$的新形式仍然是逻辑回归所使用的形式吗？推导$P(y|x)$的新形式来证明你的答案。

3. 高斯贝叶斯分类器和逻辑回归

现在，考虑我们的高斯贝叶斯分类器的以下假设(不是“朴素”的)：

- y是符合伯努利分布的布尔变量，参数$\pi=P(y=1)$，$P(y=0)=1-\pi$.
- $x=[x_1, x_2]^T$，即每个样本只考虑两个特征，每个特征为连续随机变量。假设$P(x_1, x_2|y=k)$是一个二元高斯分布 $N(\mu_{1k}, \mu_{2k}, \sigma_1, \sigma_2, \rho)$，其中$\mu_{1k}$和$\mu_{2k}$是$x_1$和$x_2$ 的均值，σ_1 和σ_2 是x_1和x_2的标准差，ρ是x_1和x_2 的相关性。二元高斯分布的概率密度为：

$$P\left(x_1, x_2 \mid y=k\right)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{\sigma_2^2(x_1-\mu_{1k})^2+\sigma_1^2(x_2-\mu_{2k})^2-2\rho\sigma_1\sigma_2(x_1-\mu_{1k})(x_2-\mu_{2k})}{2(1-\rho^2)\sigma_1^2\sigma_2^2}\right]$$

问：这种“不那么朴素”的高斯贝叶斯分类器所隐含的$P(y|x)$的形式仍然是逻辑回归所使用的形式吗?推导$P(y|x)$的形式来证明你的答案。

五、第七章作业

1. 给定如下训练数据集，

$\boldsymbol{x}^1$=[3 3], $\boldsymbol{x}^2$=[4 3], y^1=1, y^2=1

$\boldsymbol{x}^3$=[1 1],y^3=-1

通过求解 SVM 的对偶问题来求解最大间隔的分离超平面。

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y^i y^j \alpha_i \alpha_j \left(\boldsymbol{x}^i\right)^T \boldsymbol{x}^j$$

$$s.t.\ \alpha_i \geq 0, i = 1, \ldots, N,$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0.$$

例题

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y^i y^j \langle x^i, x^j \rangle$$

$$\text{s.t.} \quad \alpha_i \geq 0 \quad , i=1,2,\cdots,N$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0$$

$x^1=[3,3]^T \quad y^1=1$

$x^2=[4,3]^T \quad y^2=1$

$x^3=[1,1]^T \quad y^3=-1$

$\langle x^1, x^1 \rangle = 18, \quad \langle x^1, x^2 \rangle = 21 \quad \langle x^1, x^3 \rangle = 6$

$\langle x^2, x^2 \rangle = 25 \quad \langle x^2, x^3 \rangle = 7 \quad \langle x^3, x^3 \rangle = 2$

对 $\theta_D(\alpha_1,\alpha_2) = \alpha_1+\alpha_2+\alpha_3-[9\alpha_1^2+21\alpha_1\alpha_2-6\alpha_1\alpha_3-7\alpha_2\alpha_3+\frac{25}{2}\alpha_2^2+\alpha_3^2]$

由 $\alpha_1+\alpha_2-\alpha_3=0 \Rightarrow \alpha_3=\alpha_1+\alpha_2$

代入$\theta_D(\alpha_1,\alpha_2)$得

$$\theta_D(\alpha_1,\alpha_2) = -4\alpha_1^2-10\alpha_1\alpha_2-6\frac{1}{2}\alpha_2^2+2\alpha_1+2\alpha_2$$

令 $\dfrac{\partial\theta_D(\alpha_1,\alpha_2)}{\partial\alpha_1} = -8\alpha_1-10\alpha_2+2=0 \Rightarrow \alpha_1=\frac{1}{4}-\frac{5}{4}\alpha_2$

例题

代入$\theta_D(\alpha_1\ \alpha_2)$得

$\cancel{\theta_D(\alpha_1})$ $\theta_D(\alpha_1,\alpha_2)=-\frac{1}{4}\alpha_2^2-\frac{1}{2}\alpha_2+\frac{1}{4}$

易知$\cancel{\theta_D}$当$\alpha_2=-1$时θ_D取极大值。但由于$\alpha_2\geq 0$，所以$\alpha_2=0$

时θ_D取极大值.

即$\alpha_2=0$，$\alpha_1=\frac{1}{4}-\frac{5}{4}\alpha_2=\frac{1}{4}$，$\alpha_3=\alpha_1+\alpha_2=\frac{1}{4}$

$w^*=\frac{1}{4}x^1-\frac{1}{4}x^3=[\frac{1}{2},\frac{1}{2}]^T$ $b^*=y^1-(w^*)^T x^1=-2$

线性可分SVM-作业

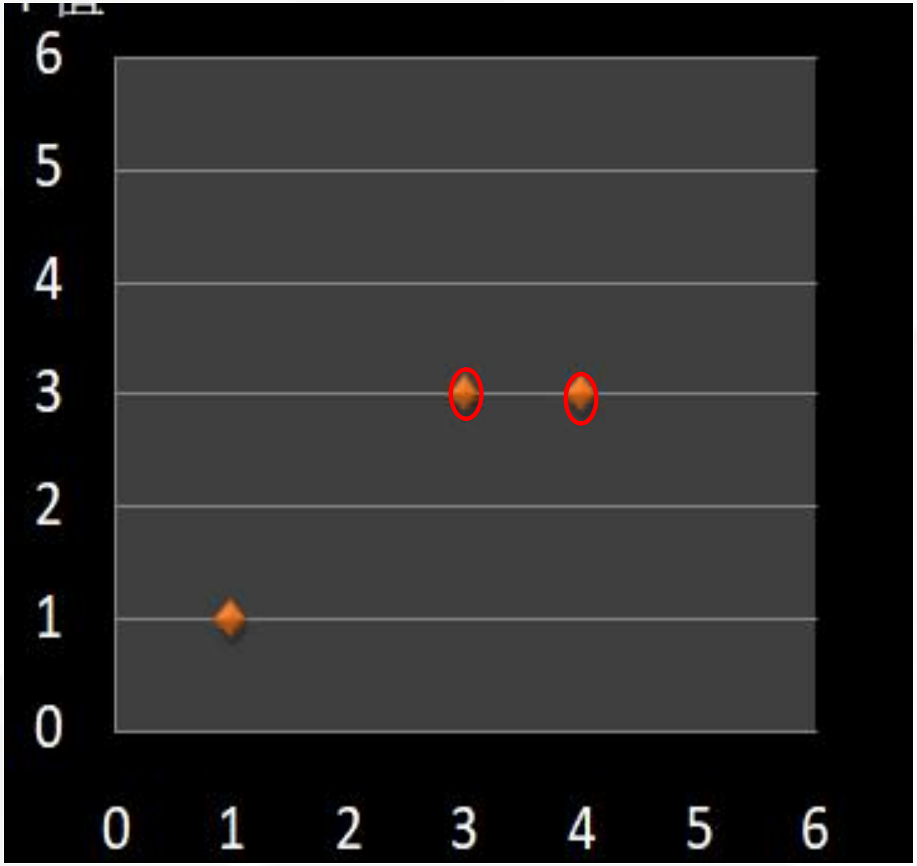
- 训练数据

  $\boldsymbol{x}^1=[3\ 3]^T$, $\boldsymbol{x}^2=[4\ 3]^T$,

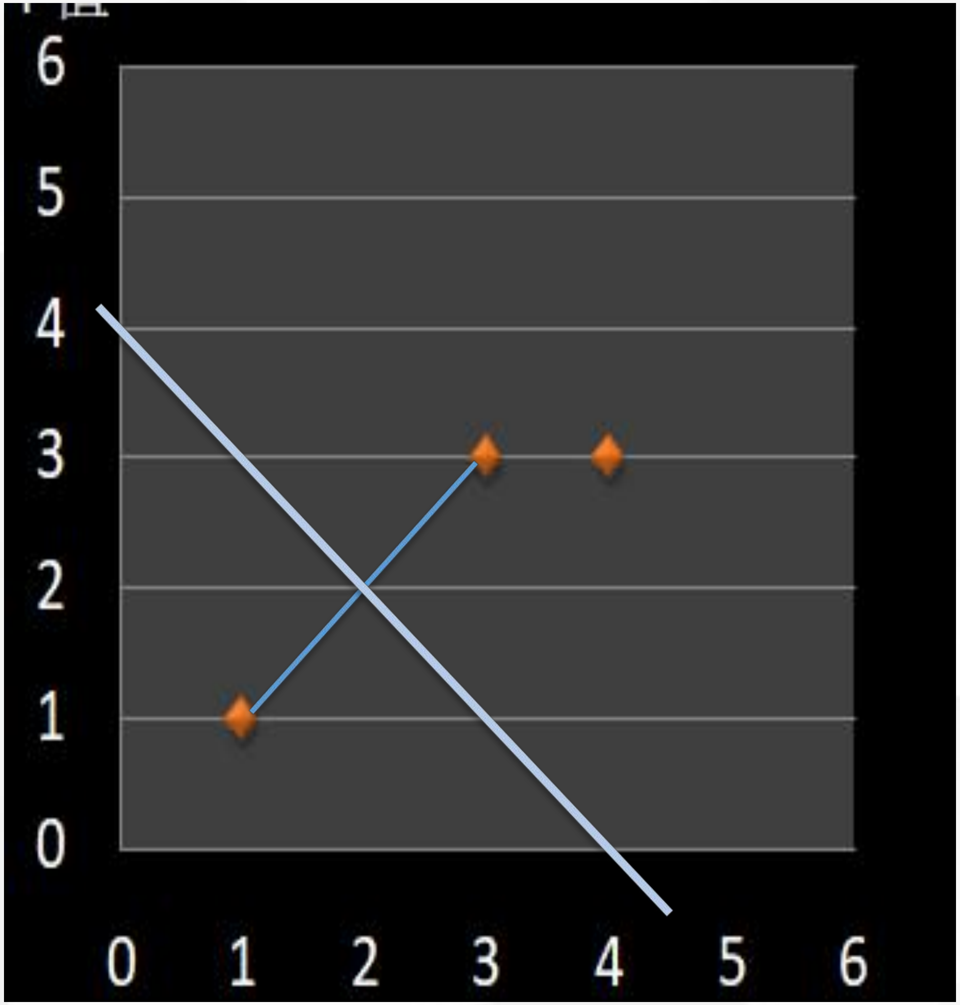
  $y^1=1$, $y^2=1$

  $\boldsymbol{x}^3=[1\ 1]$, $y^3=-1$

  通过求解对偶问题来得到最大间隔分类器，写出分类超平面及判别函数。



作业题



- 法向量平行于：$\boldsymbol{x}^1-\boldsymbol{x}^3$
- 过$\boldsymbol{x}^1$, $\boldsymbol{x}^3$连线的中点

- 可写出分离超平面对应的方程：$\boldsymbol{w}=[2,2]^T$, $b=-8$

- y=-1这一类只有一个样本$\boldsymbol{x}^3$，所以肯定在决策边界上。$(\boldsymbol{w}^T\boldsymbol{x}^3-b)y^3=4$,所以对$\boldsymbol{w}$, b缩放1/4即可得到$\boldsymbol{w}^*$, b^*;

- $\boldsymbol{w}^*=[1/2,1/2]^T$, $b^*=-2$