模式识别作业六

数据科学与计算机学院　17大数据与人工智能

17341015　陈鸿峥

问题 1 (§5 Q4). 考虑判别中用的超平面。

(a) 证明在从超平面$g(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0 = 0$到点$\mathbf{x}_a$的距离为$|g(\mathbf{x}_a)| / \|\mathbf{w}\|$，且对应的点是约束条件$g(\mathbf{x}) = 0$下的满足使$\|\mathbf{x} - \mathbf{x}_a\|^2$最小的$\mathbf{x}$。

(b) 证明$\mathbf{x}_a$到超平面的投影为

$$\mathbf{x}_p = \mathbf{x}_a - \frac{g(\mathbf{x}_a)}{\|\mathbf{w}\|^2}\mathbf{w}$$

解答. (a) 优化问题如下

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x} - \mathbf{x}_a\|^2 \\ \text{s.t.} \quad & g(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0 = 0 \end{aligned}$$

构造拉格朗日函数（这里为方便计算，在约束项前面添加了常数2）

$$\begin{aligned} f(\mathbf{x}, \lambda) &= \|\mathbf{x} - \mathbf{x}_a\|^2 + 2\lambda[g(\mathbf{x})] \\ &= \|\mathbf{x} - \mathbf{x}_a\|^2 + 2\lambda\left[\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0\right] \\ &= (\mathbf{x} - \mathbf{x}_a)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_a) + 2\lambda\left(\mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0\right) \\ &= \mathbf{x}^{\mathrm{T}}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\mathbf{x}_a + \mathbf{x}_a^{\mathrm{T}}\mathbf{x}_a + 2\lambda\left(\mathbf{x}^{\mathrm{T}}\mathbf{w} + w_0\right) \end{aligned}$$

分别求偏导有

$$\begin{cases} \frac{\partial f(\mathbf{x},\lambda)}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{x}_a + \lambda\mathbf{w} = 0 \\ \frac{\partial f(\mathbf{x},\lambda)}{\partial \lambda} = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0 = 0 \end{cases}$$

联立可解得

$$\lambda = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}_a + w_0}{\mathbf{w}^{\mathrm{T}}\mathbf{w}}$$

进而有

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_a - \lambda\mathbf{w} \\ &= \begin{cases} \mathbf{x}_a - \left[\frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}_a + w_0}{\mathbf{w}^{\mathrm{T}}\mathbf{w}}\right]\mathbf{w} & \mathbf{w} \neq \mathbf{0} \\ \mathbf{x}_a & \mathbf{w} = \mathbf{0} \end{cases} \end{aligned}$$

代入原式得到距离最小值

$$\begin{aligned}\|\mathbf{x}-\mathbf{x}_a\| &= \left\|\mathbf{x}_a - \left[\frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}_a + w_0}{\mathbf{w}^{\mathrm{T}}\mathbf{w}}\right]\mathbf{w} - \mathbf{x}_a\right\| \\ &= \left\|\left(\frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}_a + w_0}{\mathbf{w}^{\mathrm{T}}\mathbf{w}}\right)\mathbf{w}\right\| \\ &= \frac{|g(\mathbf{x}_a)|\,\|\mathbf{w}\|}{\|\mathbf{w}\|^2} = \frac{|g(\mathbf{x}_a)|}{\|\mathbf{w}\|}\end{aligned}$$

符合题意

(b) 由于(a)已经得到从超平面到$\mathbf{x}$的最小距离，而$\mathbf{x}_a$的投影即为拉格朗日求导等于零求出来的$\mathbf{x}$值，即

$$\begin{aligned}\mathbf{x}_p &= \mathbf{x}_a - \lambda\mathbf{w} \\ &= \mathbf{x}_a - \frac{g(\mathbf{x}_a)}{\|\mathbf{w}\|^2}\mathbf{w}\end{aligned}$$

其中

$$\lambda = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}_a + w_0}{\mathbf{w}^{\mathrm{T}}\mathbf{w}} = \frac{g(\mathbf{x}_a)}{\|\mathbf{w}\|^2}$$

问题 2 (§5 Q14). *考虑平方误差和准则函数（式(43)）*

$$J_s(\mathbf{a}) = \sum_{i=1}^{n}(\mathbf{a}^{\mathrm{T}}\mathbf{y}_i - b_i)^2$$

令$b_i = b$，取如下6个训练点：

$$\begin{array}{llll}\omega_1 & : (1,5)^{\mathrm{T}} & (2,9)^{\mathrm{T}} & (-5,-3)^{\mathrm{T}} \\ \omega_2 & : (2,-3)^{\mathrm{T}} & (-1,-4)^{\mathrm{T}} & (0,2)^{\mathrm{T}}\end{array}$$

(a) 计算它的Hessian矩阵

(b) 假定二次准则函数，计算最优学习率η

解答. 将训练集数据ω_1和ω_2构成矩阵，并扩增一个维度做规范化，得到

$$Y = \begin{pmatrix} 1 & 1 & 5 \\ 1 & 2 & 9 \\ 1 & -5 & -3 \\ -1 & -2 & 3 \\ -1 & 1 & 4 \\ -1 & 0 & -2 \end{pmatrix} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b \\ b \\ b \\ b \\ b \\ b \end{pmatrix}$$

平方误差准则函数为

$$J_s(\mathbf{a}) = \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{a}^{\mathrm{T}}\mathbf{y}_i - b\right)^2 = \frac{(Y\mathbf{a}-\mathbf{b})^{\mathrm{T}}(Y\mathbf{a}-\mathbf{b})}{2}$$

(a) 对准则函数求导得到

$$\nabla J_s(\mathbf{a}) = Y^{\mathrm{T}}(Y\mathbf{a}-\mathbf{b})$$

进而Hessian矩阵为

$$H = \nabla^2 J_s(\mathbf{a}) = Y^{\mathrm{T}}Y = \begin{pmatrix} 6 & -1 & 6 \\ -1 & 35 & 36 \\ 6 & 36 & 144 \end{pmatrix}$$

(b) 由公式(14)，最优学习率为

$$\eta = \frac{[\nabla J_s(\mathbf{a})]^{\mathrm{T}}\,\nabla J_s(\mathbf{a})}{[\nabla J_s(\mathbf{a})]^{\mathrm{T}}\,H\nabla J_s(\mathbf{a})}$$

通过求解特征方程，得到H的特征值[1]为$5.417, 24.57, 155.0$，进而由二次型的性质[2]

$$\frac{1}{\lambda_{\max}} = 0.006452 \le \eta \le 0.1846 = \frac{1}{\lambda_{\min}}$$

问题 3 (§5 Q21)**.** *证明MSE解法中的尺度因子α和Fisher线性判别（5.8.2节）的对应关系为*

$$\alpha = \left[1 + \frac{n_1 n_2}{n}\left(\mathbf{m}_1-\mathbf{m}_2\right)^{\mathrm{T}}\mathbf{S}_w^{-1}\left(\mathbf{m}_1-\mathbf{m}_2\right)\right]^{-1}$$

解答. 由等式(54)有权重向量

$$\mathbf{w} = \alpha n\mathbf{S}_w^{-1}\left(\mathbf{m}_1-\mathbf{m}_2\right)$$

其中$\mathbf{w}$满足

$$\left[\frac{1}{n}\mathbf{S}_w + \frac{n_1 n_2}{n^2}\left(\mathbf{m}_1-\mathbf{m}_2\right)\left(\mathbf{m}_1-\mathbf{m}_2\right)^{\mathrm{T}}\right]\mathbf{w} = \mathbf{m}_1-\mathbf{m}_2$$

将$\mathbf{w}$代入上式得到

$$\left[\frac{1}{n}\mathbf{S}_w + \frac{n_1 n_2}{n^2}\left(\mathbf{m}_1-\mathbf{m}_2\right)\left(\mathbf{m}_1-\mathbf{m}_2\right)^{\mathrm{T}}\right]\alpha n\mathbf{S}_w^{-1}\left(\mathbf{m}_1-\mathbf{m}_2\right) = \mathbf{m}_1-\mathbf{m}_2$$

令

$$\theta = 1 + \frac{n_1 n_2}{n^2}\left(\mathbf{m}_1-\mathbf{m}_2\right)^{\mathrm{T}}\mathbf{S}_w^{-1}\left(\mathbf{m}_1-\mathbf{m}_2\right)$$

[1]特征值可通过$|H-\lambda I|=0$求解，这里则直接使用numpy指令np.linalg.eig进行计算。

[2]二次型Q有$\min_{\|\mathbf{x}\|=1} Q(\mathbf{x}) = \lambda_{\min}$和$\max_{\|\mathbf{x}\|=1} Q(\mathbf{x}) = \lambda_{\max}$

则

$$\alpha\theta\left(\mathbf{m}_1-\mathbf{m}_2\right)=\mathbf{m}_1-\mathbf{m}_2$$

由于这条等式对于所有$\mathbf{m}_1$和$\mathbf{m}_2$成立，因此我们有$\alpha\theta=1$，或者下式成立

$$\alpha=\left[1+\frac{n_1n_2}{n^2}\left(\mathbf{m}_1-\mathbf{m}_2\right)^{\mathrm{T}}\mathbf{S}_w^{-1}\left(\mathbf{m}_1-\mathbf{m}_2\right)\right]^{-1}$$

问题 4 (§5 Q28)**.** *在5.10.2节给出的线性规划公式包含了一个极小化的人工变量τ，且满足约束条件$\mathbf{a}^{\mathrm{T}}y_i+\tau>b_i$及$\tau\geq 0$。证明得到的权向量使得以下准则函数极小化：*

$$J_\tau(\mathbf{a})=\max_{\mathbf{a}^{\mathrm{T}}\mathbf{y}_i\leq b_i}\left[b_i-\mathbf{a}^{\mathrm{T}}\mathbf{y}_i\right]$$

解答. 实际上即找

$$\forall i:\min\left\{t:t\geq 0,\mathbf{a}^{\mathrm{T}}\mathbf{y}_i+t>b_i\right\}$$

而目标是使权重向量$\mathbf{a}$最小化准则函数

$$J_t(\mathbf{a})=\max_{i:\mathbf{a}^{\mathrm{T}}\mathbf{y}_i\leq b_i}\left(b_i-\mathbf{a}^{\mathrm{T}}\mathbf{y}_i\right)$$

将其分为线性可分和线性不可分两种情况处理。

(a) 线性可分情况下，存在$\mathbf{a}_o$使得$\mathbf{a}_o^{\mathrm{T}}\mathbf{y}_i=b_i$，那么显然

$$\forall t>0,i:\ \mathbf{a}_o^{\mathrm{T}}\mathbf{y}_i+t>b_i$$

因此我们有对于所有的i

$$\begin{aligned}0&\leq\min\left\{t:t\geq 0,\mathbf{a}^{\mathrm{T}}\mathbf{y}_i+t>b_i\right\}\\&\leq\min\left\{t:t\geq 0,\mathbf{a}^{\mathrm{T}}\mathbf{y}_i+t>b_i\right\}=0\end{aligned}$$

进而

$$\min\left\{t:t\geq 0,\mathbf{a}^{\mathrm{T}}\mathbf{y}_i+t>b_i\right\}=0$$

得到的权重向量即为$\mathbf{a}_o$。由$J_t(\mathbf{a})\geq 0,\forall\mathbf{a}$且$J_t(\mathbf{a}_o)=0$，可知

$$\arg\min_{\mathbf{a}}J_t(\mathbf{a})=\mathbf{a}_o$$

这说明使$\mathbf{a}$最小化$J_t(\mathbf{a})$等价于解决修改后的问题。

(b) 线性不可分下，不存在$\mathbf{a}_o$使得$\mathbf{a}_o^{\mathrm{T}}\mathbf{y}_i = b_i$，那么

$$\begin{aligned}
\forall i:\ \min_{t,\mathbf{a}} \left\{t : t \geq 0, \mathbf{a}^{\mathrm{T}}\mathbf{y}_i + t > b_i\right\} &= \min_{t,\mathbf{a}} \left\{t : t \geq 0, t > b_i - \mathbf{a}^{\mathrm{T}}\mathbf{y}_i\right\} \\
&= \min_{t,\mathbf{a}} \left\{t : t \geq 0, t > \max_i \left(b_i - \mathbf{a}^{\mathrm{T}}\mathbf{y}_i\right)\right\} \\
&= \min_{t,\mathbf{a}} \left\{t : t > 0, t > \max_i \left(b_i - \mathbf{a}^{\mathrm{T}}\mathbf{y}_i\right)\right\} \\
&= \min_{t,\mathbf{a}} \left\{t : t > \max_{i:\mathbf{a}^{\mathrm{T}}\mathbf{y}_i \leq b_i} \left(b_i - \mathbf{a}^{\mathrm{T}}\mathbf{y}_i\right)\right\} \\
&= \min_{t,\mathbf{a}} \left\{\max_{i:\mathbf{a}^{\mathrm{T}}\mathbf{y}_i \leq b_i} \left(b_i - \mathbf{a}^{\mathrm{T}}\mathbf{y}_i\right)\right\} \\
&= \min_{\mathbf{a}} J_t(\mathbf{a})
\end{aligned}$$

问题 5 (§5 Q32). *考虑支持向量机和分属两类的训练样本：*

$$\begin{array}{llll}
\omega_1\text{:} & (1,1)^{\mathrm{T}} & (2,2)^{\mathrm{T}} & (2,0)^{\mathrm{T}} \\
\omega_2\text{:} & (0,0)^{\mathrm{T}} & (1,0)^{\mathrm{T}} & (0,1)^{\mathrm{T}}
\end{array}$$

(a) 在图中作出这*6*个训练点，构造具有最优超平面和最优间隔的权向量。

(b) 哪些是支持向量？

(c) 通过寻找拉格朗日待定系数α_i来构造在对偶空间的解，并将它与*(a)*中的结果比较。

解答. (a) 按序编号

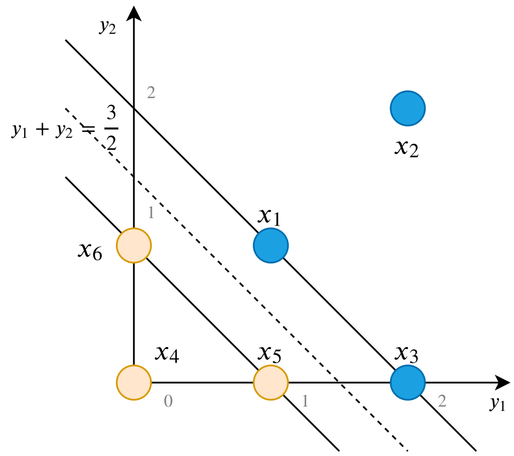
$$\omega_1 : \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$\omega_2 : \mathbf{x}_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{x}_6 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

且$z_1 = z_2 = z_3 = -1, z_4 = z_5 = z_6 = 1$。如下图所示，考虑做恒等映射$\mathbf{y}_k = \varphi(\mathbf{x}_k) = \mathbf{x}_k$，则在$\mathbf{y}$空间上最优超平面为$y_1 + y_2 = 3/2$，即

$$(3/2, -1, -1)^{\mathrm{T}}\,(1, y_1, y_2) = 0$$

将其做放缩变换$(\times 2)$，得到权向量为$(3, -2, -2)^{\mathrm{T}}$。



最优间隔为从样本点到超平面的最小距离为$1/2\sin(\pi/4)=\sqrt{2}/4$。

(b) 从上图可以清晰看出，支持向量为

$$\{\mathbf{x}_1,\mathbf{x}_3,\mathbf{x}_5,\mathbf{x}_6\}=\left\{(1,1)^{\mathrm{T}},(2,0)^{\mathrm{T}},(1,0)^{\mathrm{T}},(0,1)^{\mathrm{T}}\right\}$$

(c) 最大化式(109)给出的准则函数

$$L(\boldsymbol{\alpha})=\sum_{k=1}^{n}\alpha_k-\frac{1}{2}\sum_{k,j}^{n}\alpha_k\alpha_j z_k z_j \mathbf{y}_j^{\mathrm{T}}\mathbf{y}_k$$
$$\text{s.t.}\sum_{k=1}^{n}z_k\alpha_k=0,\ \alpha_k\geq 0$$

使用约束条件，可用下式消除一个元

$$\alpha_6=\alpha_1+\alpha_2+\alpha_3-\alpha_4-\alpha_5$$

求偏导得到

$$\begin{bmatrix}-1&-2&-2&0&1\\-2&-5&2&-1&1\\-2&2&-5&1&1\\0&-1&1&-1&-1\\1&1&3&-1&-2\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\\\alpha_3\\\alpha_4\\\alpha_5\end{bmatrix}=\begin{bmatrix}-2\\-2\\-2\\0\\0\end{bmatrix}$$

但是这个方程组不相容，因此最大值一定在边界处取到（即有些α_i消失了）。接下来尝试分别令每一个$\alpha_i=0$来求解。

- $\alpha_1 = 0$时，

$$\frac{\partial L\left(0, \alpha_2, \alpha_3, \alpha_4, \alpha_5\right)}{\partial \alpha_i} = 0$$

可解得$\boldsymbol{\alpha} = (0, -1/10, -1/10, 8/5, -8/5, -4/5)^{\mathrm{T}}$

- $\alpha_2 = 0$时，下式导致不相容

$$\frac{\partial L\left(\alpha_1, 0, \alpha_3, \alpha_4, \alpha_5\right)}{\partial \alpha_i} = 0$$

- $\alpha_3 = 0$时，下式导致不相容

$$\frac{\partial L\left(\alpha_1, \alpha_2, 0, \alpha_4, \alpha_5\right)}{\partial \alpha_i} = 0$$

- $\alpha_4 = 0$时，

$$\frac{\partial L\left(\alpha_1, \alpha_2, \alpha_3, 0, \alpha_5\right)}{\partial \alpha_i} = 0$$

解得$\boldsymbol{\alpha} = 1/5(16, 0, 4, 0, 14, 6)^{\mathrm{T}}$，满足$\alpha_i \geq 0$的限制，此时准则函数$L(\boldsymbol{\alpha}) = 4$

- $\alpha_5 = 0$时，

$$\frac{\partial L\left(\alpha_1, \alpha_2, \alpha_3, \alpha_4, 0\right)}{\partial \alpha_i} = 0$$

得到$\boldsymbol{\alpha} = 1/5(2, 2, 2, 0, 0, 6)^{\mathrm{T}}$，满足$\alpha_i \geq 0$的限制，此时准则函数$L(\boldsymbol{\alpha}) = 1.2$

故$\boldsymbol{\alpha} = 1/5(16, 0, 4, 0, 14, 6)^{\mathrm{T}}$时，准则函数$L$取得最大值，且满足限制条件。

接下来计算权向量$\mathbf{a}$。视$\mathbf{a}$为变量，尝试最小化式(108)中的

$$L(\mathbf{a}, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{a}\|^2 - \sum_{k=1}^{n} \alpha_k \left[z_k \mathbf{a}^{\mathrm{T}} \mathbf{y}_k - 1\right]$$

求导有

$$\frac{\partial L}{\partial \mathbf{a}} = \mathbf{a} - \sum_{k=1}^{n} \alpha_k z_k \mathbf{y}_k = \mathbf{0}$$

进而$\mathbf{a} = (a_0, -2, -2)^{\mathrm{T}}$。

最后确定a_0。使用支持向量$\mathbf{y} = (1, 1, 1)^{\mathrm{T}}$，满足$\mathbf{a}^{\mathrm{T}} \mathbf{y}_1 z_1 = 1$，得到

$$-\begin{pmatrix} a_0 \\ -2 \\ -2 \end{pmatrix} (1, 1, 1) = -a_0 + 4 = 1$$

求得$a_0 = 3$，最终的权向量为$\mathbf{a} = (3, -2, -2)^{\mathrm{T}}$。