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

解答 (§5 Q4). (a) 构造拉格朗日函数

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

分别求偏导有

$$\begin{aligned} \tfrac{\partial f(\mathbf{x},\lambda)}{\partial \mathbf{x}} &= \mathbf{x}-\mathbf{x}_a+\lambda\mathbf{w} = 0 \\ \tfrac{\partial f(\mathbf{x},\lambda)}{\partial \lambda} &= \mathbf{w}^t\mathbf{x}+w_0 = 0 \end{aligned}$$

联立可解得

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

进而有

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

最终得到距离最小值

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

(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

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

其中

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

解答 (§5 Q14). 将训练集数据ω_1和ω_2构成矩阵，并扩增一个维度，得到

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

平方误差准则函数为

$$\mathbf{J}_s(\mathbf{a}) = \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{a}^t\mathbf{y}_i - b\right)^2 = \frac{(\mathbf{Ya}-\mathbf{b})^t(\mathbf{Ya}-\mathbf{b})}{2}$$

(a) 对准则函数求导得到

$$\nabla J_s(\mathbf{a}) = Y^t(\mathbf{Ya}-\mathbf{b}) = \begin{pmatrix} 6a_1 & - & a_2 & + & 6a_3 & + & 0 \\ -a_1 & + & 35a_2 & + & 36a_3 & + & 3b \\ 6a_1 & + & 36a_2 & + & 144a_3 & - & 16b \end{pmatrix}$$

进而Hessian矩阵为

$$\mathbf{H} = \mathbf{Y}^t\mathbf{Y} = \begin{pmatrix} 6 & -1 & 6 \\ -1 & 35 & 36 \\ 6 & 36 & 144 \end{pmatrix}$$

(b) 最优学习率

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

通过求解特征方程，得到$\mathbf{H}$的特征值为$5.417, 24.57, 155.0$，进而$0.006452 \le \eta \le 0.1846$

解答 (§5 Q21). 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

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

将$\mathbf{w}$代入得到

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

进而

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

其中

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

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

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

解答 (§5 Q28)**.** 实际上即找

$$\min\left\{t: t\geq 0, \mathbf{a}^t\mathbf{y}_i+t>b_i \text{ for all } i\right\}$$

而目标是使权重向量$\mathbf{a}$最小化准则函数

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

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$\mathbf{a}_o^t\mathbf{y}_i=b_i$，那么显然

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

因此我们有对于所有的i

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

进而

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

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$\mathbf{a}_o^t\mathbf{y}_i = b_i$，那么

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

解答 (§5 Q32). (a) 按序编号

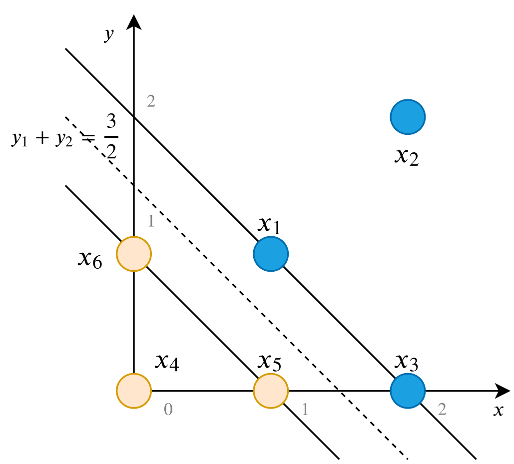
$$
\omega_1 : \mathrm{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathrm{x}_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \mathrm{x}_3 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}
$$

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

且$z_1 = z_2 = z_3 = -1, z_4 = z_5 = z_6 = 1$。最优超平面为$y_1 + y_2 = 3/2$或

$$
(3/2, -1, -1)^t (1, y_1, y_2) = 0
$$

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



最优间隔为从样本点到超平面的最小距离，如上图所示，为$\sqrt{2}/4$。

(b) 从上图可以清晰看出，支持向量为

$$\{\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_5, \mathbf{x}_6\} = \{(1,1)^t, (2,0)^t, (1,0)^t, (0,1)^t\}$$

(c) 最大化式(109)给出的准则函数

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

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，

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

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

- $\alpha_2 = 0$时，下式导致不相容

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

- $\alpha_3 = 0$时，下式导致不相容

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

- $\alpha_4 = 0$时，

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

解得$\boldsymbol{\alpha}=1/5(16,0,4,0,14,6)^t$，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$\boldsymbol{\alpha}=1/5(2,2,2,0,0,6)^t$，满足$\alpha_i\geq 0$的限制，此时准则函数$L(\boldsymbol{\alpha})=1.2$

故$\boldsymbol{\alpha}=1/5(16,0,4,0,14,6)^t$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

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

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$\mathbf{a}=(0,-2,-2)^t$。

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

$$-\begin{pmatrix}0\\-2\\-2\end{pmatrix}(1,1,1)=-a_0+4=1$$

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