hw9

1. **题目叙述: 已知正例点 x_1=(1,2)^T , x_2=(2,3)^T , x_3=(3,3)^T , 负例点 x_4=(2,1)^T , x_5=(3,2)^T , 试求最大间隔分离超平面和分类决策函数, 并在图上画出分离超平面, 间隔边界及支持向量**

$$f = \max \; \alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5 - \frac{1}{2}\left(5\alpha_1^2+13\alpha_2^2+18\alpha_3^2+5\alpha_4^2+15\alpha_5^2\right.$$
$$-8\alpha_1\alpha_2+9\alpha_1\alpha_3-4\alpha_1\alpha_4-7\alpha_1\alpha_5+15\alpha_2\alpha_3-7\alpha_2\alpha_4-12\alpha_2\alpha_5-9\alpha_3\alpha_4$$
$$\left.-15\alpha_3\alpha_5+8\alpha_4\alpha_5\right)$$

把 $\alpha_1+\alpha_2+\alpha_3-\alpha_4=\alpha_5$ 代入得.

$$f=-2\alpha_1^2-2\alpha_2^2-\frac{1}{2}\alpha_3^2-\alpha_4^2+2\alpha_1+2\alpha_2+2\alpha_3-2\alpha_1\alpha_2-3\alpha_1\alpha_4-\alpha_2\alpha_3-\alpha_3\alpha_4$$

求此最大值.

易知 $\alpha_4=0$.

$$f=-2\alpha_1^2-2\alpha_2^2-\frac{1}{2}\alpha_3^2+2\alpha_1+2\alpha_2+2\alpha_3-2\alpha_1\alpha_2-\alpha_2\alpha_3$$

求极值

$$\begin{cases}4\alpha_1+2\alpha_2=2.\\ 4\alpha_2+2\alpha_1+\alpha_3=2\\ \alpha_2+\alpha_3=2\end{cases} \Rightarrow \begin{array}{l}\alpha_1=0.75\\ \alpha_2=-0.5\\ \alpha_3=2.5\end{array}$$

违反KKT. $\lambda_i \geq 0$.

$-4\alpha_1-\alpha_3=0$

$4\alpha_1+\alpha_3=2$

$-4\times 2+2$.

故至少有1个0.

设 $\alpha_1=0$ $f=-2\alpha_2^2-\frac{1}{2}\alpha_3^2+2\alpha_2+2\alpha_3-\alpha_2\alpha_3$

$\alpha_2=0$ $f=-2\alpha_1^2-\frac{1}{2}\alpha_3^2+2\alpha_1+2\alpha_3$

...

$\alpha_3=0$

$\alpha_2=0$ $\alpha_3=2$. $f=2$.

$\alpha_1=\frac{1}{2}$ $\alpha_3=2$ $f=\frac{5}{2}$

$f=1$

故 $\alpha_2=0$ $\alpha_1=\frac{1}{2}$ $\alpha_3=2$.

$\alpha_4=0$ $\alpha_5=\frac{5}{2}$.

① 因此 支持向量为 x_1, x_3, x_5.

② $w=\sum_{i=1}^{n}\alpha_i y_i x_i$

$$=\frac{1}{2}\begin{pmatrix}1\\2\end{pmatrix}+2\begin{pmatrix}3\\3\end{pmatrix}-\frac{5}{2}\begin{pmatrix}3\\2\end{pmatrix}$$

$$=\begin{pmatrix}-1\\2\end{pmatrix}.$$

$\text{margin}=\frac{2}{\|w\|}=\frac{2}{\sqrt{5}}$.

③ $\cancel{3+b=5}$.

$\cancel{b=\pm 2}$

代入(1,2)得.

$3+b=1$

$b=-2$.

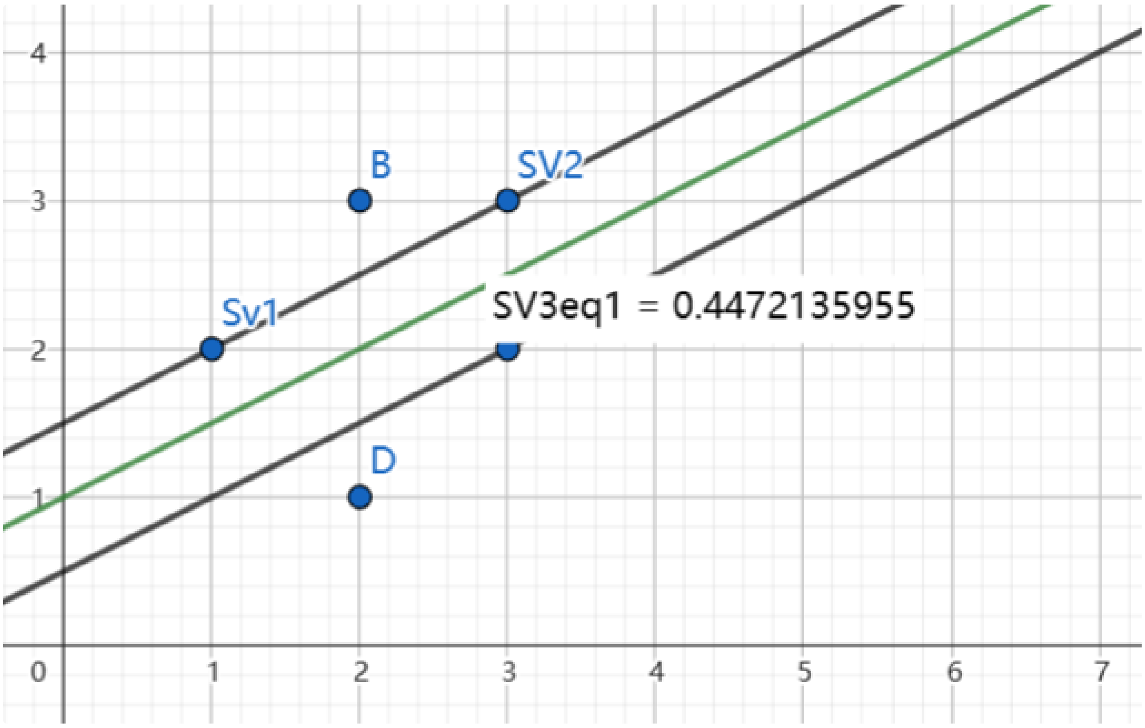
超平面为 $\begin{pmatrix}-1\\2\end{pmatrix}^T x-2=0$.

即 $-x_1+2y-2=0$.

④ 分类决策函数为

$\text{sign}(-x_1+2y-2)$.

超平面图为



2. 证明

- 计算 $\frac{\partial}{\partial w_j} L_{CE}(\mathbf{w}, b)$，其中

  $$L_{CE}(\mathbf{w}, b) = -[y \log \sigma(\mathbf{w} \cdot \mathbf{x} + b) + (1-y) \log(1 - \sigma(\mathbf{w} \cdot \mathbf{x} + b))]$$

  为 Logistic Regression 的 Loss Function。

  - 已知

$$\frac{\partial}{\partial z}\sigma(z) = \frac{\partial}{\partial z}\frac{1}{1+e^{-z}} = -\left(\frac{1}{1+e^{-z}}\right)^2 \times (-e^{-z})$$
$$= \sigma^2(z)\left(\frac{1-\sigma(z)}{\sigma(z)}\right) = \sigma(z)(1-\sigma(z))$$

$$-y_i x_j + y_i x_j \sigma(m) - \sigma(m) x_j + y_i \sigma(m) x_j$$

令 $w^T x + b = m$.

$$L_{CE} = -y \log \sigma(w^T x + b) + (1-y) \log [1 - \sigma(w^T x + b)]$$

$$\frac{\partial L_{CE}}{\partial w_j} = \cancel{-y \log} \; -y \frac{1}{\sigma(m)} x_j \sigma(m)[1-\sigma(m)] + (1-y) \frac{1}{1-\sigma(m)} - \sigma(m)[1-\sigma(m)] x_j$$

化简得 $\frac{\partial L_{CE}}{\partial w_j} = -y x_j (1-\sigma(m)) - (1-y) x_j [1-\sigma(m)]$.