人工智能基础作业 9

傅申 PB20000051

1.

根据所给数据, 对偶问题为:

$$
\begin{aligned}
\min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)-\sum_{i=1}^{N}\alpha_i \\
& =\frac{5}{2}\alpha_1^2+\frac{13}{2}\alpha_2^2+9\alpha_3^2+\frac{5}{2}\alpha_4^2+\frac{13}{2}\alpha_5^2+8\alpha_1\alpha_2+9\alpha_1\alpha_3-4\alpha_1\alpha_4-7\alpha_1\alpha_5+15\alpha_2\alpha_3 \\
& -7\alpha_2\alpha_4-12\alpha_2\alpha_5-9\alpha_3\alpha_4-15\alpha_3\alpha_5+8\alpha_4\alpha_5-(\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5) \\
\text{s.t.} \quad & \alpha_1+\alpha_2+\alpha_3-\alpha_4-\alpha_5=0 \\
& \alpha_i\geq 0, i=1,2,3,4,5
\end{aligned}
$$

将 $\alpha_1+\alpha_2+\alpha_3-\alpha_4=\alpha_5$ 代入目标函数, 得到:

$$
\begin{aligned}
\ell(\boldsymbol{\alpha}) &= 2\alpha_1^2+\alpha_2^2+\frac{1}{2}\alpha_3^2+\alpha_4^2+2\alpha_1\alpha_2-2\alpha_1\alpha_4+\alpha_2\alpha_3+\alpha_3\alpha_4-2(\alpha_1+\alpha_2+\alpha_3) \\
&= \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\begin{pmatrix} 4 & 2 & 0 & -2 \\ 2 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ -2 & 0 & 1 & 2 \end{pmatrix}\boldsymbol{\alpha}-2(1\ \ 1\ \ 1\ \ 0)\boldsymbol{\alpha}
\end{aligned}
$$

对 $\boldsymbol{\alpha}$ 求导, 得到:

$$
\frac{\partial \ell}{\partial \boldsymbol{\alpha}}=\begin{pmatrix} 4 & 2 & 0 & -2 \\ 2 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ -2 & 0 & 1 & 2 \end{pmatrix}\boldsymbol{\alpha}-\begin{pmatrix} 2 \\ 2 \\ 2 \\ 0 \end{pmatrix}=\begin{pmatrix} 4\alpha_1+2\alpha_2-2\alpha_4-2 \\ 2\alpha_1+2\alpha_2+\alpha_3-2 \\ \alpha_2+\alpha_3+\alpha_4-2 \\ -2\alpha_1+\alpha_3+2\alpha_4 \end{pmatrix}
$$

令导数为 0, 方程无解, 故只能在边界上取到最小值:

- 当 $\alpha_1=0$ 时, 最小值 $\ell(0,0,2,0)=-2$;
- 当 $\alpha_1\neq 0,\alpha_2=0$ 时, 最小值 $\ell\left(\frac{1}{2},0,2,0\right)=-\frac{5}{2}$;
- 当 $\alpha_1\neq 0,\alpha_2\neq 0,\alpha_3=0$ 时, 最小值 $\ell\left(\frac{1}{2},\frac{1}{2},0,\frac{1}{2}\right)=-1$;
- 当 $\alpha_1\neq 0,\alpha_2\neq 0,\alpha_3\neq 0,\alpha_4=0$ 时, 无最小值.

因此, $\ell(\boldsymbol{\alpha})$ 在 $\boldsymbol{\alpha}^*=\left(\frac{1}{2},0,2,0\right)^{\mathrm{T}}$ 时达到最小, 此时 $\alpha_5=\alpha_1+\alpha_2+\alpha_3-\alpha_4=\frac{5}{2}$, 说明实例点 $\boldsymbol{x}_1$, $\boldsymbol{x}_3$, $\boldsymbol{x}_5$ 为支持向量. 求得最优化问题的解 $\boldsymbol{w}^*$ 和 b^* 为

$$
\boldsymbol{w}^*=\frac{1}{2}\boldsymbol{x}_1+2\boldsymbol{x}_3-\frac{5}{2}\boldsymbol{x}_5=\begin{pmatrix} -1 \\ 2 \end{pmatrix}
$$

$$
b^*=1-\frac{1}{2}\times 5-2\times 9+\frac{5}{2}\times 7=-2
$$

因此, SVM 的最大间隔分离超平面为

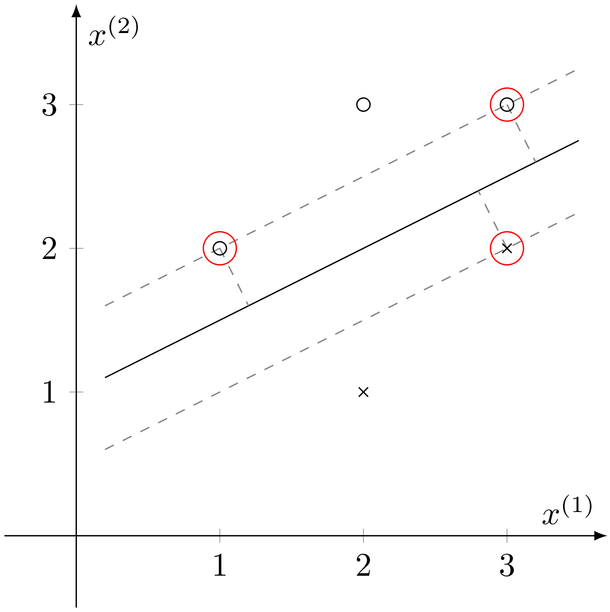
$$
-x^{(1)}+2x^{(2)}-2=0\Leftrightarrow x^{(2)}=\frac{1}{2}x^{(1)}+1
$$

分类决策函数为

$$f(\boldsymbol{x}) = \operatorname{sgn}\big(-x^{(1)} + 2x^{(2)} - 2\big)$$

由上面的分析, 作出图像如下, 其中

- 圆圈表示正例点, 叉表示负例点;
- 实线表示分离超平面, 虚线表示其间隔边界;
- 支持向量用红色圆圈圈出:



2.

首先, 有

$$\frac{\partial}{\partial z}\sigma(z) = \frac{\partial}{\partial z}\frac{1}{1+\mathrm{e}^{-z}} = \sigma(z)(1-\sigma(z))$$

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

记 $g = \boldsymbol{w}\boldsymbol{x} + b, f(g) = L_{\mathrm{CE}}(\boldsymbol{w}, b) = -[y\log(\sigma(g)) + (1-y)\log(1-\sigma(g))]$, 则有

$$\frac{\partial g}{\partial \boldsymbol{w}} = \boldsymbol{x}^{\mathrm{T}} \iff \frac{\partial g}{\partial w_j} = x_j$$

$$\begin{aligned}
\frac{\partial}{\partial g}f(g) &= -\left[y\frac{1}{\sigma(g)}\frac{\partial}{\partial g}\sigma(g) + (1-y)\frac{1}{1-\sigma(g)}\frac{\partial}{\partial g}(1-\sigma(g))\right] \\
&= -[y(1-\sigma(g)) - (1-y)\sigma(g)] \\
&= -y + \cancel{y\sigma(g)} + \sigma(g) - \cancel{y\sigma(g)} \\
&= \sigma(g) - y = \sigma(\boldsymbol{w}\boldsymbol{x} + b) - y
\end{aligned}$$

由链式法则, 有

$$\frac{\partial L_{\mathrm{CE}}}{\partial w_j} = \frac{\partial L_{\mathrm{CE}}}{\partial g}\cdot\frac{\partial g}{\partial w_j} = (\sigma(\boldsymbol{w}\boldsymbol{x} + b) - y)x_j$$