机器学习 作业一

Hugo Zhang

1

异或的数据为:

$X : \{(0,0),(0,1),(1,0),(1,1)\}$
$y : \{0,1,1,0\}$

该数据线性不可分，因此无法使用感知机模型分类。
可以使用反证法证明：假设存在感知机模型$\{w_1, w_2, b\}$能正确分类异或数据，分别带入上述四个数据，则有以下式子成立:

$$0 \cdot w_1 + 0 \cdot w_2 + b < 0$$
$$0 \cdot w_1 + 1 \cdot w_2 + b > 0$$
$$1 \cdot w_1 + 0 \cdot w_2 + b > 0$$
$$1 \cdot w_1 + 1 \cdot w_2 + b < 0$$

将第一个式子和最后一个式子相加，可得$w_1 + w_2 + 2b < 0$;
而将第二个式子和第三个式子相加，可得$w_1 + w_2 + 2b > 0$，出现矛盾。故结论得证。

2

对于问题

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i^2$$
$$s.t. y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, ..., N$$

使用拉格朗日乘子法，对第一种约束引进乘子$\alpha_i \geq 0$，第二种约束引进乘字$\beta_i \geq 0$，构造拉格朗日函数如下:

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i^2 - \sum_{i=1}^{N}\alpha_i[y_i(w \cdot x_i + b) - 1 + \xi_i] - \sum_{i=1}^{N}\beta_i\xi_i$$

对每个w, b, ξ_i求偏导为0，注意这里ξ和之前的不太一样:

$$\nabla_w L = w - \sum_{i=1}^{N}\alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{N}\alpha_i y_i x_i$$
$$\nabla_b L = -\sum_{i=1}^{N}\alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i y_i = 0$$
$$\nabla_{\xi_i} L = 2C\xi_i - \alpha_i - \beta_i = 0 \Rightarrow \alpha_i + \beta_i = 2C\xi_i$$

将w带入拉格朗日函数，并考虑以上约束，得到：

$$
\begin{aligned}
L(w,b,\xi,\alpha,\beta) &= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i\cdot x_j + \sum_{i=1}^{N}\xi_i(C\xi_i-\alpha_i-\beta_i) \\
&= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i\cdot x_j - \frac{1}{4C}\sum_{i=1}^{N}(\alpha_i+\beta_i)^2
\end{aligned}
$$

此时我们无法像之前一样消掉β，因此最后得到的对偶形式为

$$
\max_{\alpha,\beta}\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i\cdot x_j - \frac{1}{4C}\sum_{i=1}^{N}(\alpha_i+\beta_i)^2
$$

此处可知\beta只有一个约束，为了得到max应有\beta为0，从而可消去\beta

$$
s.t.\sum_{i=1}^{N}\alpha_i y_i = 0, \alpha_i \geq 0, \beta_i \geq 0, i=1,2,...,N
$$