黄磊　计702　2022E8013282156

1. 给定如下训练数据集，

x^1=[3 3]，x^2=[4 3]，y^1=1，y^2=1

x^3=[1 1], y^3=-1

通过求解 SVM 的原始问题来求解最大间隔的分离超平面。

解答：

由题：$w = [w_1, w_2]^T$

则优化问题为：

$$\min_{w} \ \frac{1}{2}\left||w|\right|^2 = \frac{1}{2}(w_1^2 + w_2^2)$$

$$s.t.\ 3w_1 + 3w_2 + b \geq 1$$

$$4w_1 + 3w_2 + b \geq 1$$

$$-w_1 - w_2 - b \geq 1$$

1 式和 3 式相加，得到：

$$w_1 + w_2 \geq 1$$

结合目标函数，解得此问题的最优解为:

$$w_1 = w_2 = \frac{1}{2}, \qquad b = -2$$

于是得到最大间隔分离超平面为：

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

其中，$x_1 = (3{,}3)^T$, $x_3 = (1{,}1)^T$ 为支持向量。

2. 给定如下训练数据集，

x^1=[3 3], x^2=[4 3], y^1=1, y^2=1

x^3=[1 1], y^3=-1

通过求解 SVM 的对偶问题来求解最大间隔的分离超平面。

解答：

对偶问题：

$$\min_{\alpha} \ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{N}\alpha_i$$

$$=\frac{1}{2}(18\alpha_1^2+25\alpha_2^2+2\alpha_3^2+42\alpha_1\alpha_2-14\alpha_2\alpha_3-12\alpha_1\alpha_3)-\alpha_1-\alpha_2-\alpha_3$$

$s.t.\ \alpha_1+\alpha_2-\alpha_3=0$

$\alpha_i\geq 0, i=1,2,3$

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

$$s(\alpha_1,\alpha_2)=4\alpha_1^2+\frac{13}{2}\alpha_2^2+10\alpha_1\alpha_2-2\alpha_1-2\alpha_2$$

求偏导数：

$$\frac{\partial s}{\partial\alpha_1}=8\alpha_1+10\alpha_2-2$$

$$\frac{\partial s}{\partial\alpha_2}=13\alpha_2+10\alpha_1-2$$

取偏导数为0，得到极值点为：

$$(\alpha_1,\alpha_2)=\left(\frac{3}{2},-1\right)$$

但不满足正值条件，因此极值在边界取得。

当 $\alpha_1=0$，最小值为 $s\left(0,\frac{2}{13}\right)=-\frac{2}{13}$；当 $\alpha_2=0$，最小值为 $s\left(\frac{1}{4},0\right)=-\frac{1}{4}$；

因此，最小值点为：

$$(\alpha_1,\alpha_2)=\left(\frac{1}{4},0\right),\qquad \alpha_3=\alpha_1+\alpha_2=\frac{1}{4}$$

因此得到：

$$w_1^*=w_2^*=\frac{1}{2},\qquad b^*=-2$$

分离超平面为：

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

3. 推导软间隔SVM的对偶形式。

解答：对于原问题：

$$\min_{w} \ \frac{1}{2}\left|\left|w\right|\right|^2$$

$$s.t. \quad y_i(w\cdot x_i+b)-1\geq 0$$

引入拉格朗日乘子：

$$\alpha_i\geq 0\,, i=1,2,3,\dots,N$$

得到拉格朗日函数：

$$L(w,b,\alpha)=\frac{1}{2}\left|\left|w\right|\right|^2-\sum_{i=1}^{N}\alpha_i y_i(w\cdot x_i+b)+\sum_{i=1}^{N}\alpha_i$$

根据对偶性，原问题的对偶问题是极大极小问题：

$$\max_{\alpha} \ \min_{w,b} L(w,b,\alpha)$$

先利用 L 对 w,b 求偏导：

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{N}\alpha_i y_i x_i=0$$

$$\frac{\partial L}{\partial b}=\sum_{i=1}^{N}\alpha_i y_i=0$$

解得：

$$w=\sum_{i=1}^{N}\alpha_i y_i x_i$$

$$\sum_{i=1}^{N}\alpha_i y_i=0$$

带入 L 得到：

$$\min_{w,b} \ L(w,b,\alpha)=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)+\sum_{i=1}^{N}\alpha_i$$

之后求对 α 得极大，得到对偶问题：

$$\min_{\alpha} \ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{N}\alpha_i$$

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

$$\alpha_i\geq 0, i=1,2,\dots,N$$

4. Show that, irrespective of the dimensionality of the data space, a data set consisting of just two data points (call them x(1) and x(2), one from each class) is sufficient to determine the maximum-margin hyperplane. Fully explain your answer, including giving an explicit formula for the solution to the hard margin SVM (i.e., w) as a function of x(1)and x(2).

假设数据空间维数为N，得到数据点对应为：

$$x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \dots, x_N^{(1)}\right),\ y^{(1)} = -1$$

$$x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, \dots, x_N^{(2)}\right),\ y^{(2)} = 1$$

求解SVM原问题为：

$$\min_{w,b} \frac{1}{2}\left||w|\right|^2 = \frac{1}{2}(w_1^2 + w_2^2 + \cdots + w_N^2)$$

$$s.t. -\left(wx^{(1)} + b\right) \geq 1$$

$$\left(wx^{(2)} + b\right) \geq 1$$

将两个不等式相加得到：

$$w\left(x^{(2)} - x^{(1)}\right) \geq 2$$

另 $\alpha_i = x_i^{(2)} - x_i^{(1)}$，则上式展开即得到：

$$\sum_{i=1}^{N} w_i \alpha_i \geq 2$$

根据柯西不等式：

$$(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_N^2)(w_1^2 + w_2^2 + \cdots + w_N^2) \geq (w_1\alpha_1 + w_2\alpha_2 + \cdots + w_N\alpha_N)^2$$

因此：

$$\frac{1}{2}(w_1^2 + w_2^2 + \cdots + w_N^2) \geq \frac{1}{2}\frac{(w_1\alpha_1 + w_2\alpha_2 + \cdots + w_N\alpha_N)^2}{(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_N^2)}$$

$$= \frac{2}{(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_N^2)}$$

当且仅当：

$$\frac{w_1}{\alpha_1} = \frac{w_2}{\alpha_2} = \cdots = \frac{w_N}{\alpha_N} = \frac{2}{(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_N^2)}$$

时候取等号，此时得到最优解，进而确定了最优分类面。

5. Gaussian kernel takes the form:

$$K(\boldsymbol{x}, z) = \exp\left(-\frac{\|\boldsymbol{x}-z\|^2}{2}\right)$$

Try to show that the Gaussian kernel can be expressed as the inner product of an infinite-dimensional feature vector.

Hint: Making use of the following expansion, and then expanding the middle factor as a power series.

$$K(\boldsymbol{x}, z) = \exp\left(-\frac{\boldsymbol{x}^T\boldsymbol{x}}{2\sigma^2}\right)\exp\left(-\frac{\boldsymbol{x}^T z}{2\sigma^2}\right)\exp\left(-\frac{z^T z}{2\sigma^2}\right)$$

解答：

首先，有泰勒级数展开：

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + R(n) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

对于高斯核函数：

$$K(\boldsymbol{x}, \boldsymbol{z}) = \exp\left\{-\frac{||\boldsymbol{x}-\boldsymbol{z}||^2}{2}\right\}$$

$$= \exp\left\{-\frac{1}{2}(\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{z} + \boldsymbol{z}^T\boldsymbol{z})\right\}$$

$$= \exp\left\{-\frac{1}{2}||\boldsymbol{x}||^2\right\}\exp\{\boldsymbol{x}^T\boldsymbol{z}\}\exp\left\{-\frac{1}{2}||\boldsymbol{z}||^2\right\}$$

左右两项都是常数，将其乘积结果记作 C ，即：

$$C = \exp\left\{-\frac{1}{2}||\boldsymbol{x}||^2 - \frac{1}{2}||\boldsymbol{z}||^2\right\}$$

因此得到高斯核为：

$$K(\boldsymbol{x}, \boldsymbol{z}) = \exp\left\{-\frac{||\boldsymbol{x}-\boldsymbol{z}||^2}{2}\right\}$$

$$= C\mathrm{e}^{\boldsymbol{x}^T\boldsymbol{z}}$$

$$= C\sum_{k=0}^{\infty}\frac{(\boldsymbol{x}^T\boldsymbol{z})^k}{k!}$$

$$= C\sum_{k=0}^{\infty}\frac{(\sum_{i=1}^{N} x_i z_i)^k}{k!}$$

对于分子使用多项式展开，系数和是 k ，多项式展开结果为：

$$\left(\sum_{i=1}^{N} x_i z_i\right)^k = \sum_{l=1}^{L} \frac{k!}{k_{l1}!\,k_{l2}!\dots k_{lN}!}(x_1 z_1)^{k_{l1}}(x_2 z_2)^{k_{l2}}\dots(x_N z_N)^{k_{lN}}$$

其中：

$$\sum_{i=1}^{N} k_{li} = k, \qquad L = \frac{(k+N-1)!}{k!\,(N-1)!}$$

因此高斯核变形为：

$$K(\boldsymbol{x},\boldsymbol{z}) = \exp\left\{-\frac{\left|\left|\boldsymbol{x}-\boldsymbol{z}\right|\right|^2}{2}\right\}$$

$$= C\sum_{k=0}^{\infty} \frac{\left(\sum_{i=1}^{N} x_i z_i\right)^k}{k!}$$

$$= C\sum_{k=0}^{\infty} \frac{1}{k!}\sum_{l=1}^{L} \frac{k!}{k_{l1}!\,k_{l2}!\dots k_{lN}!}(x_1 z_1)^{k_{l1}}(x_2 z_2)^{k_{l2}}\dots(x_N z_N)^{k_{lN}}$$

$$= \sum_{k=0}^{\infty}\sum_{l=1}^{L} \frac{C}{k_{l1}!\,k_{l2}!\dots k_{lN}!}(x_1 z_1)^{k_{l1}}(x_2 z_2)^{k_{l2}}\dots(x_N z_N)^{k_{lN}}$$

$$= \sum_{k=0}^{\infty}\sum_{l=1}^{L} \sqrt{\frac{C}{k_{l1}!\,k_{l2}!\dots k_{lN}!}}(x_1)^{k_{l1}}(x_2)^{k_{l2}}\dots(x_N)^{k_{lN}}$$

$$\cdot\sqrt{\frac{C}{k_{l1}!\,k_{l2}!\dots k_{lN}!}}(z_1)^{k_{l1}}(z_2)^{k_{l2}}\dots(z_N)^{k_{lN}}$$

$$= \sum_{k=0}^{\infty}\sum_{l=1}^{L} \varphi_{k_l}(\boldsymbol{x})\varphi_{k_l}(\boldsymbol{z})$$

令：

$$\Phi_{\mathrm{k}}(\boldsymbol{x}) = \left[\varphi_{k_1}(\boldsymbol{x}), \varphi_{k_2}(\boldsymbol{x}), \varphi_{k_3}(\boldsymbol{x}), \dots, \varphi_{k_L}(\boldsymbol{x})\right]$$

因此得到：

$$K(\boldsymbol{x},\boldsymbol{z}) = \exp\left\{-\frac{\left|\left|\boldsymbol{x}-\boldsymbol{z}\right|\right|^2}{2}\right\} = \sum_{k=0}^{\infty} < \Phi_{\mathrm{k}}(\boldsymbol{x}), \Phi_{\mathrm{k}}(\boldsymbol{z}) >$$

故高斯核可被映射为两个无限维空间向量的内积。