十一、SVM

一、LR与SVM的区别

1、从损失函数（本质区别）：

SVM：对LR的一种改造，只关注离边境近的点计算损失函数（给0）

LR：对分类错的，无论远近，都有值进行损失，得到是一个概率（给一个很小的值）

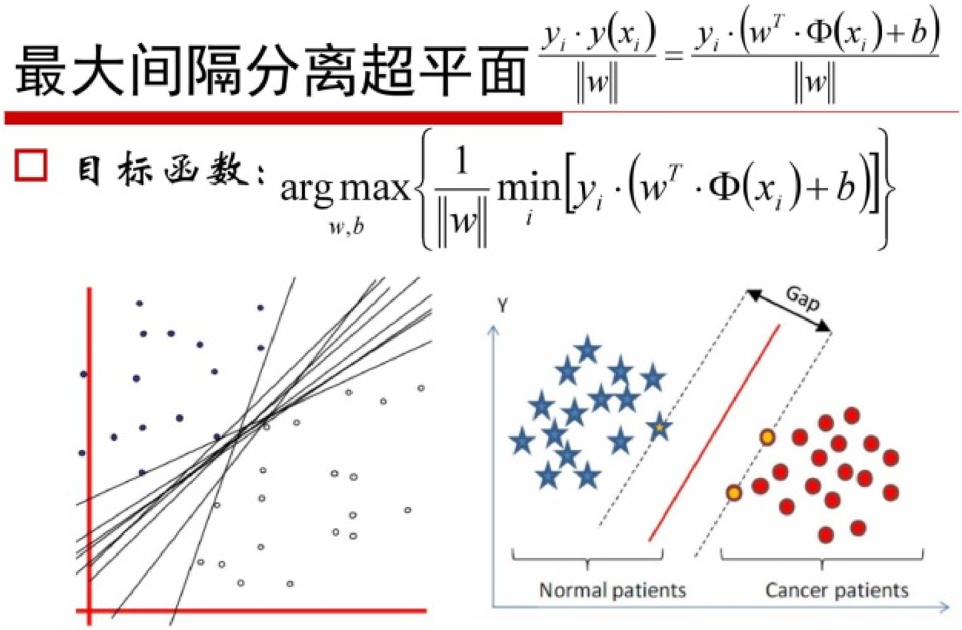
2、支持向量机是有约束的，LR是无约束，约束跟支持向量紧密相关

3、LR可解释性强

4、SVM不能给出概率结果，LR可以给出概率结果

二、SVM损失函数

SVM损失函数的就是支持向量间的最大间隔



目标函数：

$$\min_{w,b} \frac{1}{2}\|w\|^2$$

$$s.t. \quad y_i\left(w^T \cdot \Phi(x_i) + b\right) \geq 1, \quad i = 1, 2, \cdots, n$$

求解目标函数的问题实际上是一个凸二次规划问题（凸函数，二次）。为什么要转换为对偶问题，因为是有约束的最优化问题。求解问题，拉格朗日乘子法。

计算拉格朗日函数的对偶函数

$$\boxed{\begin{aligned} w &= \sum_{i=1}^{n} \alpha_i y_i \Phi(x_n) \\ 0 &= \sum_{i=1}^{n} \alpha_i y_i \end{aligned}}$$

$$\begin{aligned}
L(w,b,\alpha) &= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(w^T \cdot \Phi(x_i) + b \right) - 1 \right) \\
&= \frac{1}{2} w^T w - w^T \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) - b \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i \\
&= \frac{1}{2} w^T \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) - w^T \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) - b \cdot 0 + \sum_{i=1}^{n} \alpha_i \\
&= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \left(\sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) \right)^T \sum_{i=1}^{n} \alpha_i y_i \Phi(x_i) \\
&= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \Phi^T(x_i) \Phi(x_j)
\end{aligned}$$

$$a^* = \arg\max_{\alpha} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \Phi^T(x_i) \Phi(x_j) \right)$$

整理目标函数：添加负号

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

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

$$\alpha_i \geq 0, \quad i = 1,2,\ldots,n$$

三、线性支持向量机（不可分，泛化能力强）

线性SVM的目标函数

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} \xi_i$$

$$s.t. \quad y_i (w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1,2,\cdots,n$$

$$\xi_i \geq 0, \quad i = 1,2,\cdots,n$$

整理，得到对偶问题：

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

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

$$0 \le \alpha_i \le C, \quad i = 1,2,\ldots,n$$

四、核函数

核函数其实就是相当于特征组合，高斯核把特征映射到无线维，很有可能过拟合