SVM (Support Vector Machine)：支持向量机模型

应用：实现对数据集的线性/非线性划分、分类以及对未知样本点的预测。

目标函数：

$$J(w,b,i) = \arg\max_{\boxed{w,b}} \min(d_i)$$

$\arg\max_{w,b}$ ⇒ 分割带最宽的超平面

$\min(d_i)$ ⇒ 所有样本点到某个分割面间最小距离，也为分割带

w,b ⇒ 分割面：$w'x+b=0$，w,b 为分割面参数

分割带边界虚线上点称为支持向量，为SVM模型核心点。

△ 分割带越宽模型泛化性越好，划分样本能力与可信度越高。

此时的分割面为最优超平面。

函数间隔：$\hat{\gamma}_i = y_i \cdot (w'x_i + b)$

y_i：样本点所属类别 ⇒ 1/-1

$w'x_i+b \geq 1$ ⇒ y_i 预测为 +1

$w'x_i+b \leq -1$ ⇒ y_i 预测为 −1

因此得到：

线性可分SVM模型的函数间隔 $\hat{\gamma}_i \geq 1$，此为必要条件

几何间隔：

$$\gamma_i = \frac{\hat{\gamma}_i}{\|w\|} = \frac{y_i \times (w'x_i+b)}{\|w\|} = \frac{|w'x_i+b|}{\|w\|} = \underbrace{d_i}_{\text{点到超平面距离}}$$

超平面: $wx+b=0$

w, b 同比↑ 对其无影响，但影响 函数

间隔 $\hat{y}_i$，因此必须做单位化处理。

目标函数

SVM
- 线性可分模型: $J(w,b,i) = \arg_{w,b}\max\min(d_i) = \arg_{w,b}\max\frac{1}{\|w\|}\cdot\min(\hat{y}_i)$
- 非线性可分模型:

线性可分模型:

此前由函数间隔已知 $\hat{y}_i \geq 1 \Rightarrow \min \hat{y}_i = 1$

$$\therefore J(w,b,i) = \arg_{w,b}\max\frac{1}{\|w\|}\cdot\min(\hat{y}_i)$$

$$\Longleftrightarrow \begin{cases}\max\frac{1}{\|w\|} \\ s.t.\ \hat{y}_i \geq 1\end{cases} \Longleftrightarrow \begin{cases}\min\frac{1}{2}\|w\|^2 \\ s.t.\ \hat{y}_i \geq 1\end{cases}$$

函数间隔约束.

拉格朗日乘子法:

要求: ① 目标函数为 $\min(f(x))$ ② 有约束条件: $s.t.\ g(x) \leq 0$

原理: 利用拉格朗日对偶性将原始最优化问题转换为对偶问题，

即: $\min(f(x)) = \min_x\max_\lambda(L(x,\lambda)) = \min_x\max_\lambda(f(x)+\sum_{i=1}^{k}\lambda_i g_i(x))$

$= \max_\lambda\min_x(\ \ f(x) + \sum^{k}$

拉格朗日乘子

$\therefore$ e.g. $\min \frac{1}{2}\|w\|^2 = \max_\alpha \min_{w,b}\left(\frac{1}{2}\|w\|^2 + \sum_{i=1}^{n}\alpha_i(1-\hat{y}_i)\right)$

$= \max_\alpha \min_{w,b}\left(\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\alpha_i y_i \times \overbrace{(w^T x_i + b)}^{\hat{y}_i} + \sum_{i=1}^{n}\alpha_i\right)$

拉格朗日乘子法的目标函数

求解 w,b：对拉格朗日乘子法得到的目标函数求偏导，令导函数为0.

$$\begin{cases} \dfrac{\partial L(w,b,\alpha)}{\partial w} = 0 \Rightarrow w - \sum_{i=1}^{n}\alpha_i \cdot y_i \cdot x_i = 0 & ① \\ \dfrac{\partial L(w,b,\alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i \cdot y_i = 0 & ② \end{cases}$$

将①，②代入 $L(w,b,\alpha)$ 得：

$\min \frac{1}{2}\|w\|^2 = \max_\alpha \min_{w,b}\left(\underbrace{\frac{1}{2}\|w\|^2}_{\frac{1}{2}\cdot w^T \cdot w} - \sum_{i=1}^{n}\alpha_i y_i \times (w^T x_i + b) + \sum_{i=1}^{n}\alpha_i\right)$

$= \max_\alpha\left(-\frac{1}{2}\cdot w^T \cdot \sum_{i=1}^{n}\alpha_i \cdot y_i \cdot x_i + \sum_{i=1}^{n}\alpha_i - 0\right)$

$= \max_\alpha \cdot \left(-\frac{1}{2}\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i \cdot y_i \cdot x_i \cdot \underbrace{\alpha_j \cdot y_j \cdot x_j}_{w^T} + \sum_{i=1}^{n}\alpha_i\right)$

新目标函数

对新目标函数求解：

$$\Leftrightarrow \begin{cases} \min_{\alpha} \left(\frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \cdot y_i \cdot x_i \cdot \alpha_j \cdot y_j \cdot x_j - \sum_{i=1}^{n} \alpha_i \right) \\ s.t. \quad \sum_{i=1}^{n} \alpha_i \cdot y_i = 0 \quad \Rightarrow \text{原始目标函数求偏导得到} \\ \qquad \alpha_i \geq 0 \quad \Rightarrow \text{拉格朗日乘子约束} \\ \\ w^{*} x + b = 0 \end{cases}$$

$$\xrightarrow{\text{解得}} \begin{cases} \hat{w} = \sum_{i=1}^{n} \hat{\alpha}_i \, y_i \cdot x_i \\ \hat{b} = y_i - \sum_{i=1}^{n} \hat{\alpha}_i \, y_i \, (\underbrace{x_i \cdot x_j}_{\text{内积}}) \end{cases}$$

ith example x_i.

$$f_k^{(i)} = \text{sim}\left(x^{(i)}, l^{(k)}\right)$$

kth landmark $l^{(k)}$.

features

$$\theta^T f = \theta_0 + \theta_1 f_1 + \cdots \theta_k f_k \cdots$$

$$\geq 0 \xRightarrow{\text{预测}} y = 1$$

△ $\theta^T f$ is the hypothesis function of kernel SVM.

$$C \cdot A + B$$

$$A = \sum_{i=1}^{m} y^{(i)} \log\left(h_\theta(x^{(i)})\right) \cdots$$

Logistic Regression : $A + \lambda \cdot B$

$$B = \frac{1}{2m} \sum_{i=1}^{n} \theta_i^2$$