Support Vector Machine (支持向量机)

1. Overview of SVM

Hard-margin SVM: max margin (w,b) data is linearly separable

Soft-margin SVM: data is linearly separable and allow a little bit outlier

Kernel SVM: linearly inseparable

2. Hard margin SVM(硬间隔分类器)

Suppose the data is linear separable, then we can use the hard margin SVM, the key point is to find the largest possible geometric margin.

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$$

$$x_i \in \mathcal{X}=\mathbf{R}^n,\ y_i \in \mathcal{Y}=\{+1,-1\},\ i=1,2,\cdots,N$$

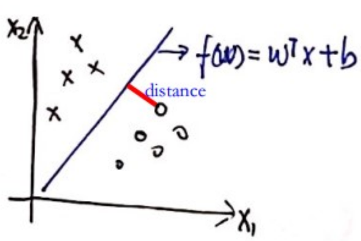
$$f(x)=\operatorname{sign}(w^* \cdot x+b^*)$$

$$\text{margin}(w,b)=\min_{\substack{w,b\\ x_i\ \text{for}\ i=1,2,\cdots,N}} \text{distance}(w,b,x_i)$$

$$=\min_{\substack{w,b,x_i\\ i=1,\cdots,N}} \frac{1}{\|w\|}|w^T x_i+b|$$

max margin

$$\Rightarrow \begin{cases} \max\limits_{w,b} \min\limits_{x_i\ i=1,\cdots,N} \frac{1}{\|w\|}|w^T x_i+b| \\ \text{such that } y_i(w^T x_i+b)>0 \Rightarrow \exists \gamma>0 \text{ s.t. } \min y_i(w^T x_i+b)=\gamma \end{cases}$$



$$\Rightarrow \max_{w,b} \min_{x_i=1,\cdots,N} \frac{1}{\|w\|} y_i(w^T x_i+b)$$

$$\Rightarrow \max_{w,b} \frac{1}{\|w\|} \min_{x_i=1,\cdots,N} y_i(w^T x_i+b) \quad \text{set } \gamma=1$$

$$\text{s.t.} \begin{cases} w^T x_i+b>0 & y_i=1 \\ w^T x_i+b<0 & y_i=-1 \end{cases}$$

$$\Rightarrow \max_{w,b} \frac{1}{\|w\|} \min_{x_i=1,\dots,N} y_i(w^T x_i + b) \quad \text{set } r=1$$

$$\Rightarrow \begin{cases} \max\limits_{w,b} \dfrac{1}{\|w\|} \\ \text{s.t. } \min y_i(w^T x + b_i) \geq 1 \\ \quad \rightarrow y_i(w^T x_i + b) \geq 1 \end{cases}$$

$$\boxed{\Rightarrow \begin{cases} \min\limits_{w,b} \|w\| = \min\limits_{w,b} \frac{1}{2} w^T w \quad \text{quadratic} \\ \text{s.that } y_i(w^T x_i + b) \geq 1 \quad \text{for } \forall i=1,\dots,N \end{cases}}$$ optimization problem !

Solve the function above, using Lagrange duality.

The Lagrange function here is:

$$\mathcal{L}(w,b,\lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \lambda_i \underbrace{(1 - y_i(w^T x_i + b))}_{\leq 0} \rightarrow$$

无约束 $\begin{cases} \min\limits_{w,b} \max\limits_{\lambda} \mathcal{L}(w,b,\lambda) \\ \text{s.t. } \lambda_i \geq 0 \end{cases}$ $\rightarrow$ primall problem

强对偶

$$\begin{cases} \max\limits_{\lambda} \min\limits_{w,b} \mathcal{L}(w,b,\lambda) \\ \text{s.t. } \lambda_i \geq 0 \end{cases}$$

凤尾 $\min\max \mathcal{L}$ $\geq$ 鸡头 $\max\min \mathcal{L}$ (弱对偶关系)

$$\min_{w,b} \mathcal{L}(w,b,\lambda) \Rightarrow \frac{\partial}{\partial b} = \frac{\partial}{\partial b}\left[\sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i y_i(w^T x_i + b)\right]$$

$$= -\sum_{i=1}^{N} y_i \lambda_i = 0 \Rightarrow \text{代入 } \mathcal{L}(w,b,\lambda)$$

$$\mathcal{L}(w,b,\lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i y_i(w^T x_i + b)$$

$$= \frac{1}{2} w^T w + \sum_{i=1}^{N}\lambda_i + \sum_{i=1}^{N}\lambda_i y_i w^T x_i + \underbrace{\sum_{i=1}^{N}\lambda_i y_i b}_{=0}$$

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{2}\cdot 2\cdot w - \sum_{i=1}^{N}\lambda_i y_i x_i \triangleq 0$$

$$\Rightarrow \boxed{w^* = \sum_{i=1}^{N}\lambda_i y_i x_i}$$

$$\mathcal{L}(w,b,\lambda) = \frac{1}{2}\left(\sum_{i=1}^{N}\lambda_i y_i x_i\right)^T\left(\sum_{i=1}^{N}\lambda_i y_i x_i\right) - \sum_{i=1}^{N}\lambda_i y_i\left(\sum_{j=1}^{N}\lambda_j y_j x_j\right)^T x_i + \sum_{i=1}^{N}\lambda_i$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i^T x_j + \sum_{i=1}^{N}\lambda_i$$

$$\begin{cases} \max\limits_{\lambda} -\frac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i^T x_j + \sum\limits_{i=1}^{N}\lambda_i \\ \text{s.t. } \lambda_i \geq 0 \quad \text{where } \lambda = \begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\end{pmatrix} \end{cases}$$

Moreover, lambda, w, b satisfies Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{cases} \frac{\partial f}{\partial w}=0 \quad \frac{\partial f}{\partial b}=0 \\ \lambda_i(1-y_i(w^T x_i+b))=0 \rightarrow \text{slackness complementary 松弛互补条件} \\ \lambda_i \geqslant 0 \\ 1-y_i(w^T x_i-b) \leqslant 0 \end{cases}$$

原问题，对偶问题具有强对偶关系 $\overset{\text{充要}}{\Leftrightarrow}$ 满足KKT条件

由此求出 w^*, b^*

$$\boxed{w^* = \sum_{i=0}^{N} \lambda_i y_i x_i}$$

$$\exists (x_k, y_k) \text{ s.t. } 1-y_k(w^T x_k+b)=0$$

$$y_k^2(w^T x_k+b)=y_k \quad y_k^2=1$$

$$\boxed{b^* = y_k - w^T x_k = y_k - \sum_{i=0}^{N} \lambda_i y_i x_i^T x_k}$$

$$f(x)=\operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i (x \cdot x_i)+b^*\right)$$

3. Soft margin SVM (软间隔分类器)

Allow some outliers, but most of the data is linear separable.

$$\min \frac{1}{2} w^T w + loss$$

$$\underline{\text{hinge loss}} \quad loss = \max\{0, 1-y_i(w^T x_i+b)\} \rightarrow \begin{array}{ll} \text{if } y_i(w^T x_i+b) \geqslant 1 & loss=0 \\ \text{if } y_i(w^T x_i+b)<1 & loss=1-y_i(w^T x_i+b) \end{array}$$

$$\text{Soft-Margin SVM: } \begin{cases} \min_{w,b} \frac{1}{2} w^T w + \max\{0, 1-y_i(w^T x_i+b)\} \\ \text{s.t. } y_i(w^T x_i+b) \geqslant 1 \end{cases}$$

$$\xi_i = 1-y_i(w^T x_i+b) \quad \xi_i \geqslant 0$$

$$\boxed{\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & y_i(w \cdot x_i+b) \geqslant 1-\xi_i, \quad i=1,2,\cdots,N \\ & \xi_i \geqslant 0, \quad i=1,2,\cdots,N \end{aligned}}$$

Same procedure as the hard margin SVM

$$L(w,b,\xi,\alpha,\mu) \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i (y_i(w \cdot x_i+b)-1+\xi_i) - \sum_{i=1}^{N} \mu_i \xi_i \quad \alpha_i \geqslant 0, \mu_i \geqslant 0。$$

Similarly，

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

$$\begin{aligned}
\max_{\alpha} \quad & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i \\
\text{s.t.} \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\
& C - \alpha_i - \mu_i = 0 \\
& \alpha_i \geqslant 0 \\
& \mu_i \geqslant 0, \quad i = 1, 2, \cdots, N
\end{aligned}$$

4. Kernel SVM (核分类器/非线性分类)

Suppose the data is not linearly separable. This method is essentially kernel function+ hard margin Svm.

The kernel function here is

$$K(x,z) = \phi(x) \cdot \phi(z)$$

positive definite kernel function:

$K: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$\forall x, z \in \mathcal{X}$, 有 $K(x,z)$ 如果 $\exists\, \phi: \mathcal{X} \to \mathbb{R}$, $\phi \in \mathcal{H}$ s.t. $K(x,z) = \langle \phi(x) \cdot \phi(z) \rangle$

那么则称 $K(x,z)$ 为正定核函数

$\Leftrightarrow$ def: $K: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$\forall x, z \in \mathcal{X}$, 有 $K(x,z)$, if K satisfy: ① symetric ② postive definite such that $K(x,z)$ is a positive definite kernel function

① Symmetric : $K(x,z) = K(z,x)$

② positive definite: 任取 $x_1, \cdots, x_N \in \mathcal{X}$, 对应的 $\underbrace{\text{Gram matrix}}_{K = [K(x_i, x_j)]}$ is 半正定的

Used in SVM：

$$W(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N}\alpha_i$$

$$\begin{aligned}
f(x) &= \operatorname{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i \phi(x_i) \cdot \phi(x) + b^*\right) \\
&= \operatorname{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i K(x_i, x) + b^*\right)
\end{aligned}$$