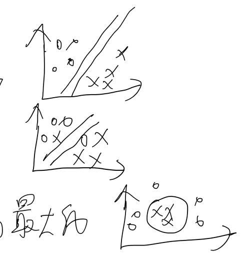
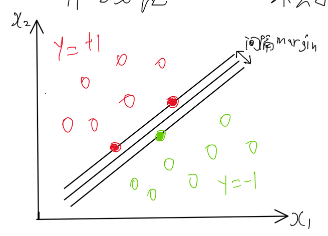
# svm

2020年8月7日 星期五 下午12:46

- 线性可分 SVM：硬间隔最大化
- 近似线性可分：软间隔最大化
- 非线性：核技巧 软间隔最大化






$x^T w + b = 0$

$margin = \frac{2}{\|w\|}$

使margin最大 $\Longleftrightarrow \max\limits_{w,b} \frac{2}{\|w\|} \Longleftrightarrow \max\limits_{w,b} \frac{4}{\|w\|^2} \Longleftrightarrow \min\limits_{w,b} \frac{1}{2}\|w\|^2$

s.t. $Y_i(X_i^T w + b) \geq 1$ (n个不等式约束) ] 原问题

在所有样本点都位于2个超平面两侧的条件下，

找到间隔最大的超平面

构建拉格朗日函数 $L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [Y_i(X_i^T w + b) - 1]$

在约束条件下，$\max\limits_{\alpha} L(w,b,\alpha) = \frac{1}{2}\|w\|^2$

∴原问题等价于 $\min\limits_{w,b} \max\limits_{\alpha} L(w,b,\alpha)$

由拉格朗日对偶性，其对偶问题为 $\max\limits_{\alpha} \min\limits_{w,b} L(w,b,\alpha)$

求 $\min\limits_{w,b} L(w,b,\alpha)$：$\nabla_w L(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i Y_i X_i = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i Y_i X_i$

$\nabla_b L(w,b,\alpha) = -\sum_{i=1}^{n} \alpha_i Y_i = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i Y_i = 0$

代入 $L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}(\alpha_i Y_i X_i^T w + \alpha_i Y_i b - \alpha_i)$

$= \frac{1}{2}\sum_{i=1}^{n} \alpha_i Y_i X_i^T \sum_{j=1}^{n} \alpha_j Y_j X_j - \sum_{i=1}^{n} \alpha_i Y_i X_i^T \sum_{j=1}^{n} \alpha_j Y_j X_j - b\sum_{i=1}^{n} \alpha_i Y_i + \sum_{i=1}^{n} \alpha_i$

$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} Y_i Y_j \alpha_i \alpha_j X_i^T X_j$

求 $\max\limits_{\alpha} \min\limits_{w,b} L(w,b,\alpha)$：$\max\limits_{\alpha}(\quad) = \min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j Y_i Y_j X_i^T X_j - \sum_{i=1}^{n} \alpha_i$ ,

s.t. $\begin{cases} \sum_{i=1}^{n} \alpha_i Y_i = 0 \\ \alpha_i \geq 0, \ i = 1,2,\cdots,n \end{cases}$

由slater条件知，存在 $\hat{w}, \hat{b}, \hat{\alpha}$，使 $\hat{w}, \hat{b}$ 是原问题的解，$\hat{\alpha}$ 是对偶问题的解

由序列最小优化 SMO 算法可得最优解 $\hat{\alpha}$

$\hat{w} = \sum_{i=1}^{n} \hat{\alpha}_i Y_i X_i = \sum_{i\in SV} \hat{\alpha}_i Y_i X_i$

由KKT条件 $\begin{cases} 乘子非负：\alpha_i \geq 0 \\ 约束条件：Y_i(X_i^T w + b) - 1 \geq 0 \\ 互补条件：\alpha_i(Y_i(X_i^T w + b) - 1) = 0 \end{cases}$ ,

∴ $\alpha_j = 0$ (非支持向量) 或 $\alpha_j > 0$ (支持向量), $Y_j(X_j^T \hat{w} + \hat{b}) - 1 = 0$

$\hat{b} = \frac{1}{Y_j} - X_j^T \hat{w} = Y_j - \sum_{i=1}^{n} \hat{\alpha}_i Y_i X_j^T X_i = Y_j - \sum_{i\in SV} \hat{\alpha}_i Y_i X_j^T X_i$

软间隔最大化：允许少量样本不满足约束 $Y_i(X_i^T w + b) \geq 1$

hinge损失 $l_{hinge}(z) = \max(0, 1-z)$

优化目标：$\min\limits_{w,b} \underbrace{\frac{1}{2}\|w\|^2}_{结构风险} + \underbrace{C}_{惩罚参数} \sum_{i=1}^{n} \underbrace{\max(0, 1 - Y_i(X_i^T w + b))}_{Loss}$

$= \min\limits_{w,b,\varepsilon} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \underbrace{\varepsilon_i}_{松弛变量}$ , s.t. $Y_i(X_i^T w + b) \geq 1 - \varepsilon_i$

$\varepsilon_i \geq 0, \ i = 1,2,\cdots,n$ ] 原问题

结构风险 经验风险

$L(w,b,\varepsilon,\alpha,\beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \varepsilon_i - \sum_{i=1}^{n} \alpha_i [Y_i(X_i^T w + b) - 1 + \varepsilon_i] - \sum_{i=1}^{n} \beta_i \varepsilon_i$

求 $\max\limits_{\alpha,\beta} \min\limits_{w,b,\varepsilon} L(w,b,\varepsilon,\alpha,\beta)$

求 $\min\limits_{w,b,\varepsilon} L(w,b,\varepsilon,\alpha,\beta)$：$\nabla_w L(w,b,\varepsilon,\alpha,\beta) = w - \sum_{i=1}^{n} \alpha_i Y_i X_i = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i Y_i X_i$

$\nabla_b L(w,b,\varepsilon,\alpha,\beta) = -\sum_{i=1}^{n} \alpha_i Y_i = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i Y_i = 0$

$\nabla_{\varepsilon_i} L(w,b,\varepsilon,\alpha,\beta) = C - (\alpha_i + \beta_i) = 0 \Rightarrow C = (\alpha_i + \beta_i)$

代入得 $\min\limits_{w,b,\varepsilon} L(w,b,\varepsilon,\alpha,\beta) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j Y_i Y_j X_i^T X_j + \underbrace{\sum_{i=1}^{n} \varepsilon_i (C - \alpha_i - \beta_i)}_{0}$

求 $\max\limits_{\alpha,\beta} \min\limits_{w,b,\varepsilon} L(w,b,\varepsilon,\alpha,\beta)$：$\max\limits_{\alpha,\beta}(\quad) = \min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j Y_i Y_j X_i^T X_j - \sum_{i=1}^{n} \alpha_i$ ,

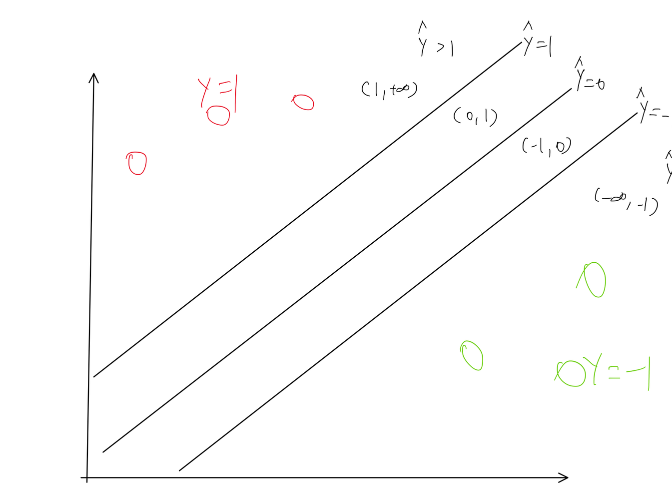
s.t. $\begin{cases} \sum_{i=1}^{n} \alpha_i Y_i = 0 \\ 0 \leq \alpha_i \leq C, \ i = 1,2,\cdots,n \end{cases}$ ( $\beta_i = C - \alpha_i \geq 0, \alpha_i \leq C$ )

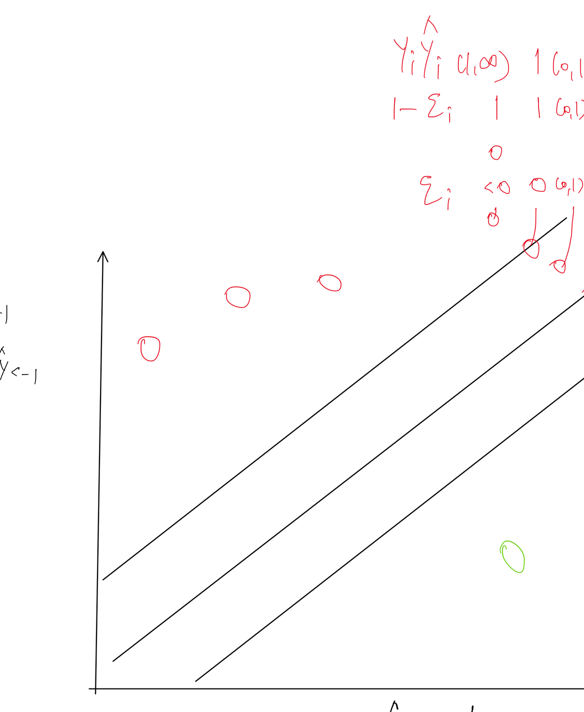
$\begin{cases} C - \alpha_i - \beta_i = 0 \\ \alpha_i \geq 0 \\ \beta_i \geq 0 \end{cases}$

KKT条件 $\begin{cases} 乘子非负：\alpha_i \geq 0, \ \beta_i \geq 0 \\ 约束条件：Y_i(X_i^T w + b) - 1 + \varepsilon_i \geq 0, \ \varepsilon_i \geq 0 \\ 互补条件：\alpha_i(Y_i(X_i^T w + b) - 1 + \varepsilon_i) = 0, \ \beta_i \varepsilon_i = 0 \end{cases}$

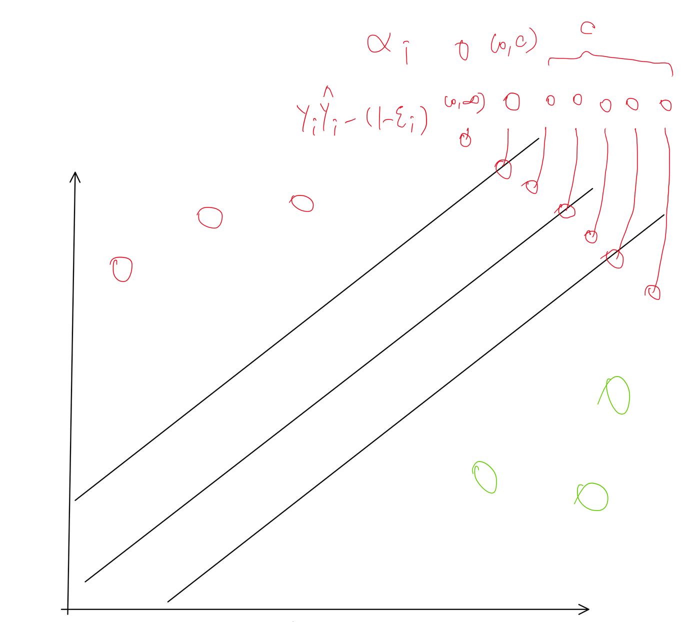
$\hat{w} = \sum_{i\in SV}^{n} \hat{\alpha}_i Y_i X_i$

$\hat{b} = \frac{1}{m}\sum_{j=1}^{m}(Y_j - \sum_{i\in SV} \hat{\alpha}_i Y_i X_j^T X_i)$，其中 $0 < \hat{\alpha}_j < C$ （对b的解不唯一，因此取所有符合条件的样本点上的平均值）

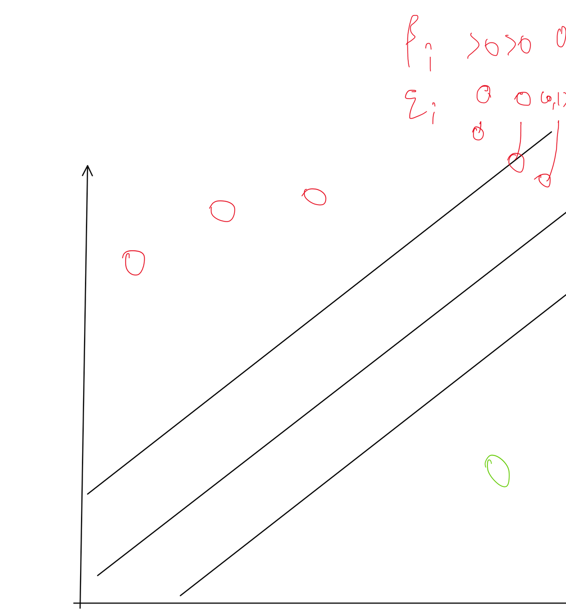



$\varepsilon_i = \max(0, 1 - Y_i\hat{Y}_i) = Loss_{hinge}$

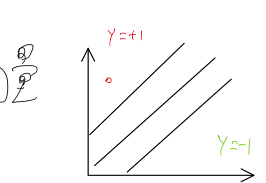



$\alpha_i : Y_i\hat{Y}_i - 1 + \varepsilon_i \geq 0$

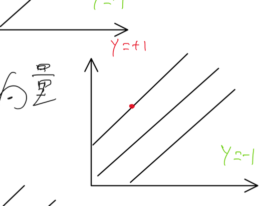



$\beta_i : \varepsilon_i \geq 0$

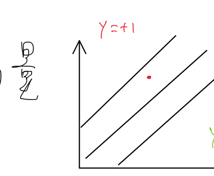
- $\hat{\alpha}_i = 0, \ \varepsilon_i = 0$：非支持向量



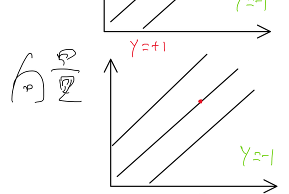

- $0 < \hat{\alpha}_i < C, \ \varepsilon_i = 0$：在间隔边界上的支持向量



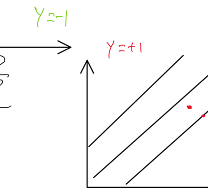

- $\hat{\alpha}_i = C, \ 0 < \varepsilon_i < 1$：支持向量




- $\hat{\alpha}_i = C, \ \varepsilon_i = 1$：支持向量




- $\hat{\alpha}_i = C, \ \varepsilon_i > 1$：被误分类的支持向量




线性可分支持向量机学习算法：

①构造并求解约束最优化问题

$\min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle - \sum_{i=1}^{n} \alpha_i$

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

$\alpha_i \geq 0, \ i = 1,2,\cdots,n$

求得最优解 $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_n)^T$

② 计算 $\hat{w}, \hat{b}$

$\hat{w} = \sum_{i=1}^{n} \hat{\alpha}_i Y_i x_i$

$\hat{b} = Y_j - \sum_{i=1}^{n} \hat{\alpha}_j Y_i \langle x_i \cdot x_j \rangle$

③ 求得

分离超平面 $\hat{w}x + \hat{b} = 0$

分类决策函数 $f(x) = sign(\hat{w}x + \hat{b})$

线性支持向量机学习算法：

①选择惩罚参数 $C > 0$，构造并求解凸二次规划问题

$\min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle - \sum_{i=1}^{n} \alpha_i$

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

$0 \leq \alpha_i \leq C, \ i = 1,2,\cdots,n$

求得最优解 $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_n)^T$

② 计算 $\hat{w}, \hat{b}$

$\hat{w} = \sum_{i=1}^{n} \hat{\alpha}_i Y_i x_i$

$\hat{b} = \frac{1}{m}\sum_{j=1}^{m}(Y_j - \sum_{i\in SV} \hat{\alpha}_i Y_i X_j^T X_i)$

③ 求得

分离超平面 $\hat{w}x + \hat{b} = 0$

分类决策函数 $f(x) = sign(\hat{w}x + \hat{b})$