Linear Support Vector Machine

1. 选取一种鲁棒性强的方法——在linear regression上做手脚
2. largest margin:

$$\begin{aligned} \max_{\mathbf{w}} \quad & \text{fatness}(\mathbf{w}) \\ \text{subject to} \quad & \mathbf{w} \text{ classifies every } (\mathbf{x}_n, y_n) \text{ correctly} \\ & \text{fatness}(\mathbf{w}) = \min_{n=1,\ldots,N} \text{distance}(\mathbf{x}_n, \mathbf{w}) \end{aligned}$$

$$\begin{aligned} \max_{\mathbf{w}} \quad & \textbf{margin}(\mathbf{w}) \\ \text{subject to} \quad & \text{every } y_n \mathbf{w}^T \mathbf{x}_n > 0 \\ & \textbf{margin}(\mathbf{w}) = \min_{n=1,\ldots,N} \text{distance}(\mathbf{x}_n, \mathbf{w}) \end{aligned}$$

3. distance(x,b,w),这里不能忽略bias

$$distance(X, b, W) = \frac{1}{||W||} |W^T x + b| \tag{1}$$

4. 数学建模

$$\begin{aligned} \max_{b,\mathbf{w}} \quad & \text{margin}(b, \mathbf{w}) \\ \text{subject to} \quad & \text{every } y_n(\mathbf{w}^T \mathbf{x}_n + b) > 0 \\ & \text{margin}(b, \mathbf{w}) = \min_{n=1,\ldots,N} \frac{1}{\|\mathbf{w}\|} y_n(\mathbf{w}^T \mathbf{x}_n + b) \end{aligned}$$

5. 由于成比例缩放的缘故：令

$$y_n(W^T X_n + b) = 1 \tag{2}$$

6. necessary constraints:

$$y_n(W^T X_n + b) \geq 1 \tag{3}$$

通过反证法可以证明(2)(3)互为充要条件，其中用到$max \frac{1}{||W||}$

$$\text{final change: max} \Longrightarrow \text{min, remove } \sqrt{\ }, \text{ add } \tfrac{1}{2}$$

$$\begin{aligned} \min_{b,\mathbf{w}} \quad & \tfrac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{subject to} \quad & y_n(\mathbf{w}^T \mathbf{x}_n + b) \geq 1 \text{ for all } n \end{aligned}$$

QP problem， 一个二次规划问题

optimal $(b, \mathbf{w}) = ?$

$$\min_{b,\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

subject to $\quad y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1,$

for $n = 1, 2, \ldots, N$

optimal $\mathbf{u} \leftarrow \text{QP}(\mathrm{Q}, \mathbf{p}, \mathrm{A}, \mathbf{c})$

$$\min_{\mathbf{u}} \quad \frac{1}{2}\mathbf{u}^T\mathrm{Q}\mathbf{u} + \mathbf{p}^T\mathbf{u}$$

subject to $\quad \mathbf{a}_m^T\mathbf{u} \geq c_m,$

for $m = 1, 2, \ldots, M$

objective function: $\mathbf{u} = \begin{bmatrix} b \\ \mathbf{w} \end{bmatrix}; \mathrm{Q} = \begin{bmatrix} 0 & \mathbf{0}_d^T \\ \mathbf{0}_d & \mathrm{I}_d \end{bmatrix}; \mathbf{p} = \mathbf{0}_{d+1}$

constraints: $\mathbf{a}_n^T = y_n \begin{bmatrix} 1 & \mathbf{x}_n^T \end{bmatrix}; c_n = 1; M = N$

SVM with QP Solver

Linear Hard-Margin SVM Algorithm

1. $\mathrm{Q} = \begin{bmatrix} 0 & \mathbf{0}_d^T \\ \mathbf{0}_d & \mathrm{I}_d \end{bmatrix}; \mathbf{p} = \mathbf{0}_{d+1}; \mathbf{a}_n^T = y_n \begin{bmatrix} 1 & \mathbf{x}_n^T \end{bmatrix}; c_n = 1$
2. $\begin{bmatrix} b \\ \mathbf{w} \end{bmatrix} \leftarrow \text{QP}(\mathrm{Q}, \mathbf{p}, \mathrm{A}, \mathbf{c})$
3. return b & $\mathbf{w}$ as g_{SVM}

- **hard-margin**: nothing violate 'fat boundary'
- **linear**: $\mathbf{x}_n$

7. 一些优势：

配合feature transformation会特别强大，鲁棒性和泛化能力都可以保证

a new possibility: non-linear SVM

	large-margin hyperplanes + numerous feature transform Φ
#	not many
boundary	sophisticated