SVM推导(Hard Magin)

要点：KKT是为了推强对偶，之后是无约束问题，不再用KKT条件求解，而是直接求导求解

数据：$(X,Y),X\in R^n,Y\in\{-1,1\}$

分类超平面：$Y=WX+b$

决策函数：$f(x_i)=sign(Wx_i+b)$

当所有的点被正确分类时：$y_i*(Wx_i+b)>0$

上式可以缩放为：$y_i*(Wx_i+b)\geq 1,min(y_i*(Wx_i+b)=1$

根据点到直线的距离公式，定义：$margin(W,b,X)=\frac{1}{||W||}|WX+b|$

优化目标：=$\arg\max\limits_{W,b}\{\frac{1}{||W||}min(y_i*(Wx_i+b))\}=\arg\max\limits_{W,b}\{\frac{1}{||W||}\}$

转化成极小值问题：公式

$$\begin{cases}min_{W,b}\frac{1}{2}W^2\\ s.t.\,y_i(WX_i+b)\geq 1\end{cases}$$

拉格朗日函数：$L(W,b,\lambda)=\frac{1}{2}W^2+\sum\limits_{i=1}^{N}\lambda_i(1-y_i(Wx_i+b))$

原问题等价于

$$\begin{cases}\underset{W,b}{min}\underset{\lambda}{max}L(W,b,\lambda)\\ s.t.\,\lambda_i\geq 0\end{cases}$$

KKT条件（最优解的必要条件）：

约束问题：$\min f(x)$

$$s.t.\,g_j(x)=0,j=1,,m\quad h_k(x)\leq 0,k=0,1,,,p$$

拉格朗日函数：$L(x,\{\lambda_j\},\{\mu_k\})=f(x)+\sum\limits_{j=1}^{m}\lambda_j g_j(x)+\sum\limits_{k=1}^{p}\mu_k h_k(x)$

KKT条件包括：

$$\begin{cases}\bigtriangledown_x L=0\\ g_j(x)=0, & j=1,2,,,m.\\ h_k(x)\leq 0, & k=1,2,,,p.\\ \mu_k\geq 0\\ \mu_k h_k(x)=0, & k=1,2,,,p.\end{cases}$$

KKT条件使用时要满足某些要求（Constraint Qualification,CQ），最常用的CQ是strict feasibility和Linearity constraint qualification:

1. 存在一个点，使得所有的约束条件都严格成立
2. 等式约束和不等式约束都是线性的

弱对偶：$\max \min f \leq \min \max f$

强对偶：$\max \min f = \min \max f$

强对偶可以推出KKT条件成立，对于凸优化问题，CQ可以推出强对偶。

显然满足这两个条件，所以KKT条件成立，强对偶定理成立。

用强对偶定理，问题继而转化成

$$\begin{cases} \underset{\lambda}{max}\underset{W,b}{min}L(W,b,\lambda) \\ s.t. \lambda_i \geq 0 \end{cases}$$

先求解，

$$\begin{cases} \underset{W,b}{min}L(W,b,\lambda) \\ s.t. \lambda_i \geq 0 \end{cases}$$

凸二次优化问题，直接求解，令$\frac{\partial L}{\partial W} = 0, \frac{\partial L}{\partial b} = 0$

得，$W = \sum\limits_{i=1}^{N} \lambda_i y_i x_i,\ \ \sum\limits_{i=1}^{N} \lambda_i y_i = 0$

带入L得，$L = \sum\limits_{i=1}^{N} \lambda_i - \frac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j$

求解b,b只与落在间隔线上的点有关，此时，必有$y_k(Wx_k + b) = 1, \lambda_k \neq 0$,

所以，$b = y_k - \sum\limits_{i=1}^{N} \lambda_i y_i x_i^T x_k$

进而求解

$$\begin{cases} \underset{\lambda}{max}L = \sum\limits_{i=1}^{N} \lambda_i - \frac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j \\ s.t. \lambda_i \geq 0,\ \sum\limits_{i=0}^{N} \lambda_i y_i = 0 \end{cases}$$

用SMO算法求解.