Problem Set 3

1.设拉格朗日函数为$\mathcal{L}(x,\lambda)=c^Tx+\lambda^T(Ax-b)$,

对应的对偶函数为$\mathcal{G}(\lambda)=inf_\lambda\ \mathcal{L}(x,\lambda)$,

而LP问题是凸问题，KTT 条件成立，满足 stationarity

$\nabla_{x^*}c^Tx+\lambda^{*T}(Ax-b)=0$

$\Longrightarrow c^T+\lambda^{*T}A=0$

以及$Ax^*-b=0$，因此该点处拉格朗日函数可以表达为

$\mathcal{L}(x^*,\lambda^*)=(-\lambda^TA)(A^{-1}b)+\lambda^T(Ax^*-b)$

$\mathcal{L}(x^*,\lambda^*)=-\lambda^Tb$

根据 Dual feasibility 得 $\lambda_i\geq 0$

LP问题的对偶问题标准形式为

$$max_\lambda\ -\lambda^Tb$$
$$s.t. \lambda\geq 0, c^T+\lambda^TA=0$$

2.

2.1证：$\omega^T\omega$是凸函数

$\iff\ ||\lambda x+(1-\lambda)y||^2\leq\lambda||x||^2+(1-\lambda)||y||$

$\iff\lambda||x||^2+(1-\lambda)||y||-(\lambda x+(1-\lambda)y)^T(\lambda x+(1-\lambda)y)\geq 0$

$\iff\lambda||x||^2+(1-\lambda)||y||-(\lambda x^T+(1-\lambda)y^T)(\lambda x+(1-\lambda)y)\geq 0$

$\iff$
$\lambda||x||^2+(1-\lambda)||y||-(\lambda^2x^Tx+\lambda(1-\lambda)(y^Tx+y^Tx)+(1-\lambda)^2y^Ty)\lambda(1-\lambda)(y^Tx+y^Tx)\geq 0$

$\iff(\lambda-\lambda^2)x^Tx+(\lambda-\lambda^2)y^Ty-\lambda(1-\lambda)(y^Tx+y^Tx)\geq 0$

$\iff(\lambda-\lambda^2)x^Tx+(\lambda-\lambda^2)y^Ty-\lambda(1-\lambda)(y^Tx+y^Tx)\geq 0$

而$\lambda\in[0,1]$，因此$\lambda\geq\lambda^2$，

$\iff x^Tx+y^Ty-(y^Tx+y^Tx)\geq 0$

$\iff(x^T-y^T)(x-y)\geq 0$

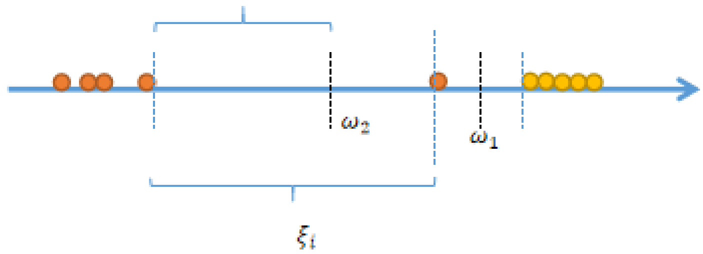
$\iff||x-y||^2\geq 0$

而$||x-y||^2\geq 0$成立，故$\omega^T\omega$是凸函数，证毕。

2.2不一定，软间隔SVM模型表达为

$$min_{\omega,b,\xi}\frac{1}{2}||\omega||^2+C\sum_{i=1}^{m}\xi_i$$
$$s.t. y^{(i)}(\omega^Tx^{(i)}+b)\geq 1-\xi_i$$
$$\xi_i\geq 0,\forall i=1,2,\ldots,m$$

考虑一维情形如下



令$\forall\xi_i = 0$，即退化为硬间隔SVM，求得决策边界为ω_1；

令$\xi_j = 0, j \neq i$，求得决策边界为ω_2；

目标函数设为f，$f(\omega_1) = \frac{1}{2}\omega_1^2$，$f(\omega_2) = \frac{1}{2}\omega_2^2 + C\xi_i$，

当$\frac{1}{2}\omega_1^2 > \frac{1}{2}\omega_2^2 + C\xi_i$时，$\xi_i$可以不为0，$\omega_2$优于$\omega_1$，因而最优解一定不是$\omega_1$.

软间隔SVM可以避免过拟合，正如上面的例子，右侧橙色点可能是噪声，用硬间隔SVM会拟合噪声；

相反，前者通过松弛变量，泛化模型，提高鲁棒性，因此某些情况下有必要使用软间隔SVM。

2.3①当$0 < \alpha_i^* < C$时，

根据KTT条件$\alpha_i^* + r_i^* = C$得$0 < r_i^* < C$，

又因为$r_i^*\xi_i^* = 0$，所以$\xi_i^* = 0$，

因为$\alpha_i^*(y^{(i)}(\omega^{*T}x^{(i)} + b^*) + \xi_i^* - 1) = 0$，

所以$y^{(i)}(\omega^{*T}x^{(i)} + b^*) + \xi_i^* - 1 = 0$，

所以$y^{(i)}(\omega^{*T}x^{(i)} + b^*) = 1$，

即 in-bound SVs 在支撑平面上。

②当$\alpha_i^* = C$时，类似的可以得到$y^{(i)}(\omega^{*T}x^{(i)} + b^*) + \xi_i^* - 1 = 0$，

而$\xi_i^* \geq 0$，因此$y^{(i)}(\omega^{*T}x^{(i)} + b^*) \leq 1$，

即 bound SVs 在支撑平面上或者在间隔内。

而往往少数的点就能确定支撑平面（n 维空间 n 个点确定一个 boundary），因此大部分的点在间隔内。