Problem Set 3

1 Lagrange Duality

> Formulate the Lagrange dual problem of the following linear programming problem
>
> $$\begin{aligned} &min \quad c^Tx \\ &s.t. \quad Ax \preceq b \end{aligned}$$
>
> where $x \in \mathbb{R}^n$ is variable, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{k \times n}$, $b \in \mathbb{R}^k$.

由题意，引进拉格朗日函数，得：

$$\mathcal{L}(x, \alpha) = c^Tx + \alpha^T(Ax - b)$$

其中，α为拉格朗日乘子，满足：$\alpha \geq 0$。

构建拉格朗日对偶问题函数：

$$\begin{aligned} \mathcal{G}(\alpha) &= \underset{x}{inf}\ \mathcal{L}(x, \alpha) \\ &= \underset{x}{inf}\ (c^Tx + \alpha^T(Ax - b)) \\ &= \underset{x}{inf}\ ((c^T + \alpha^TA)x - \alpha^Tb) \end{aligned}$$

若$c^T + \alpha^TA$不为0，则x为$+\infty$或$-\infty$，显然不合适，

因此：$c^T + \alpha^TA = 0$，对偶问题函数变为：

$$\begin{aligned} \mathcal{G}(\alpha) &= \underset{x}{inf}\ (-\alpha^Tb) \\ &= -\alpha^Tb \end{aligned}$$

再极大化对偶问题函数：

$$\underset{\alpha}{max}\ \mathcal{G}(\alpha) = \underset{\alpha}{max}\ (-\alpha^Tb)$$

综合之前提到的约束条件，得到题目中对应的拉格朗日对偶问题：

$$\begin{aligned} &\underset{\alpha}{max}\ (-\alpha^Tb) \\ &s.t. \quad \alpha \geq 0 \\ &\qquad\ \ c^T + \alpha^TA = 0 \end{aligned}$$

2 SVM

2.1 Convex Functions

> Prove $f(\omega) = \omega^T\omega$ (where $\omega \in \mathbb{R}^n$) is a convex function.

由于$\omega \in \mathbb{R}^n$， 不妨假定：

$$\omega = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

则：

$$f(\omega) = \omega^T\omega = x_1^2 + x_2^2 + \cdots + x_n^2$$

要证明$g(x)$为凸函数，即证明：

$$g(\lambda x_1 + (1-\lambda)x_2) \le \lambda g(x_1) + (1-\lambda)g(x_2)$$

令$g(x) = x^2$，有：

$$f(\omega) = g(x_1) + g(x_2) + \cdots + g(x_n)$$

下面通过数学归纳法证明$f(\omega)$为凸函数：

当$n = 1$时，显然满足凸函数定义，即$g(x) = x^2$为凸函数。

当$n > 1$时，令$\mathcal{G}(x) = g_1(x) + g_2(x)$，其中$g_1(x)$、$g_2(x)$均为凸函数，有：

$$\begin{aligned} \mathcal{G}(\lambda x_1 + (1-\lambda)x_2) &= g_1(\lambda x_1 + (1-\lambda)x_2) + g_2(\lambda x_1 + (1-\lambda)x_2) \\ &\le \lambda g_1(x_1) + (1-\lambda)g_1(x_2) + \lambda g_2(x_1) + (1-\lambda)g_2(x_2) \\ &= \lambda(g_1(x_1) + g_2(x_1)) + (1-\lambda)(g_1(x_2) + g_2(x_2)) \\ &= \lambda\mathcal{G}(x_1) + (1-\lambda)\mathcal{G}(x_2) \end{aligned}$$

因此，两个凸函数相加得到的函数仍为凸函数，故$n > 1$时，$f(\omega)$仍为凸函数。

得证：$f(\omega) = \omega^T\omega$为凸函数。

2.2 Soft-Margin for Separable Data

> Consider training a soft-margin SVM with C set to some positive constant. Suppose the training data is linearly separable. Since increasing the ξ_i can only increase the objective of the primal problem, all the training examples will have functional margin at least 1 and all the ξ_i will be equal to zero. True or false? Explain! Given a linearly separable dataset, is it necessarily better to use a hard margin SVM over a soft-margin SVM?

对于软间隔SVM，其约束条件为：

$$y^{(i)}(\omega^T x^{(i)} + b) \ge 1 - \xi_i$$

其中，ξ_i为松弛变量。

相应的优化问题为：

$$\begin{aligned} \underset{\omega,b,\xi}{min} \quad & \frac{1}{2}||\omega||^2 + C\sum_{i=1}^m \xi_i \\ s.t. \quad & y^{(i)}(\omega^T x^{(i)} + b) \ge 1 - \xi_i, \quad \forall i = 1, \ldots, m \\ & \xi_i \ge 0, \quad \forall i = 1, \ldots, m \end{aligned}$$

对应的拉格朗日对偶问题：

$$\mathcal{L}(\omega, b, \xi, \alpha, r) = \frac{1}{2}\omega^T\omega + C\sum_{i=1}^m \xi_i - \sum_{i=1}^m \alpha_i[y^{(i)}(\omega^T x^{(i)} + b) - 1 + \xi_i] - \sum_{i=1}^m r_i\xi_i$$

其中，α_i为拉格朗日乘子，满足：$\alpha_i \geq 0$，$for\ \forall\ i$。

依据KKT条件，有：

$$
\begin{aligned}
&\nabla_\omega \mathcal{L}(\omega^*, b^*, \xi^*, \alpha^*, r^*) = 0 \ \Rightarrow\ \omega^* = \sum_{i=1}^{m} \alpha_i^* y^{(i)} x^{(i)} \\
&\nabla_b \mathcal{L}(\omega^*, b^*, \xi^*, \alpha^*, r^*) = 0 \ \Rightarrow\ \sum_{i=1}^{m} \alpha_i^* y^{(i)} = 0 \\
&\nabla_{\xi_i} \mathcal{L}(\omega^*, b^*, \xi^*, \alpha^*, r^*) = 0 \ \Rightarrow\ \alpha_i^* + r_i^* = C,\ for\ \forall\ i \\
&\alpha_i^*, r_i^*, \xi_i^* \geq 0,\ for\ \forall\ i \\
&y^{(i)}(\omega^{*T} x^{(i)} + b^*) + \xi_i^* - 1 \geq 0,\ for\ \forall\ i \\
&\alpha_i^*(y^{(i)}(\omega^{*T} x^{(i)} + b^*) + \xi_i^* - 1) = 0,\ for\ \forall\ i \\
&r_i^* \xi_i^* = 0,\ for\ \forall\ i
\end{aligned}
$$

可推导出：

$$\alpha_i, r_i \geq 0,\ \alpha_i + r_i = C \ \Rightarrow\ 0 \leq \alpha_i, r_i \leq C$$

当$\alpha_i = 0$时，有：

$$\alpha_i = 0,\ \alpha_i + r_i = C \ \Rightarrow\ r_i = C$$

$$r_i = C,\ \xi_i \geq 0,\ r_i \xi_i = 0 \ \Rightarrow\ \xi_i = 0$$

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

此时，对应的训练样本不是支持向量，不会对该SVM边界的确定产生任何影响。

当$0 < \alpha_i < C$时，有：

$$0 < \alpha_i < C,\ \alpha_i + r_i = C \ \Rightarrow\ r_i \neq 0$$

$$r_i \neq 0,\ r_i \xi_i = 0 \ \Rightarrow\ \xi_i = 0$$

$$
\begin{aligned}
\alpha_i(y^{(i)}(\omega^T x^{(i)} + b) + \xi_i - 1) = 0,\ \alpha_i > 0 \ &\Rightarrow\ y^{(i)}(\omega^T x^{(i)} + b) + \xi_i - 1 = 0 \\
&\Rightarrow\ y^{(i)}(\omega^T x^{(i)} + b) = 1 - \xi_i \\
&\Rightarrow\ y^{(i)}(\omega^T x^{(i)} + b) = 1
\end{aligned}
$$

此时，对应的训练样本为支持向量，恰好在最大间隔边界上。

当$\alpha_i = C$时，对应的训练样本落在最大间隔内部或被误分类，由于题目中已告知该数据集为线性可分的，因此不存在该种情况。

综合$\alpha_i = 0$及$0 < \alpha_i < C$的两种情况，可知在该线性可分数据集上，满足：

$$
\begin{aligned}
&y^{(i)}(\omega^T x^{(i)} + b) \geq 1 \\
&\xi_i = 0,\ \forall\ i = 0
\end{aligned}
$$

而对于任意一个线性可分数据集，使用软间隔SVM的效果是要优于硬间隔SVM的。

假设线性可分的两类样本A、B之间间隔很远，但有一个A类样本十分接近B类样本所在的区域，此时若仍使用硬间隔SVM进行分类，算法会选择这一个接近B样本区域的A样本作为支持向量，此时，尽管仍能对两类样本进行分类，但分类边界会大大偏向于B类样本，出现出过拟合的现象。

2.3 In-bound Support Vectors in Soft-Margin SVMs

> Examples $x^{(i)}$ with $\alpha_i > 0$ are called support vectors (SVs). For soft-margin SVM we distinguish between *in-bound* SVs, for which $0 < \alpha_i < C$, and *bound* SVs for which $\alpha_i = C$. Show that in-bound SVs lie exactly on the margin. Argue that bound SVs can lie both on or in the margin, and that they will "usually" lie in the margin. Hint: use the KKT conditions.

由上题可知，当$0 < \alpha_i < C$，即该点为*in-bound SV*时：

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

因此，*in-bound SV*在margin上。

而当$\alpha_i = C$，即该点为*bound SV*时：

由KKT条件：

$$\begin{aligned}
&\nabla_\omega \mathcal{L}(\omega^*, b^*, \xi^*, \alpha^*, r^*) = 0 \ \Rightarrow \ \omega^* = \sum_{i=1}^{m} \alpha_i^* y^{(i)} x^{(i)} \\
&\nabla_b \mathcal{L}(\omega^*, b^*, \xi^*, \alpha^*, r^*) = 0 \ \Rightarrow \ \sum_{i=1}^{m} \alpha_i^* y^{(i)} = 0 \\
&\nabla_{\xi_i} \mathcal{L}(\omega^*, b^*, \xi^*, \alpha^*, r^*) = 0 \ \Rightarrow \ \alpha_i^* + r_i^* = C, \ for \ \forall \ i \\
&\alpha_i^*, r_i^*, \xi_i^* \geq 0, \ for \ \forall \ i \\
&y^{(i)}(\omega^{*T} x^{(i)} + b^*) + \xi_i^* - 1 \geq 0, \ for \ \forall \ i \\
&\alpha_i^*(y^{(i)}(\omega^{*T} x^{(i)} + b^*) + \xi_i^* - 1) = 0, \ for \ \forall \ i \\
&r_i^* \xi_i^* = 0, \ for \ \forall \ i
\end{aligned}$$

可得：

$$\alpha_i = C, \alpha_i + r_i = C \ \Rightarrow \ r_i = 0$$

由于$r_i = 0$，$r_i \xi_i = 0$，$\xi_i \geq 0$，且除此之外没有其他关于ξ_i的限制条件，因此：

$$\xi_i \geq 0;$$

得到：

$$y^{(i)}(\omega^T x^{(i)} + b) = 1 - \xi_i \leq 1$$

当上式等于1，即仅当$\xi_i = 0$时，该样本点位于margin上，而当上式小于1，即$\xi_i > 0$时，该样本点位于margin内，因此大多数情况下满足位于margin内的情况。

因此，*bound SV*位于margin上或margin内，且大多位于margin内。