西安交通大學

优化方法基础大作业

专业班级：

姓名学号：

联系电话：

时　间：

一、 SVM 的前世今生

（1）问题：

$$\min_{w,b,\xi_k(\forall k)} \frac{1}{2}\|w\|_2^2 + C\sum_{k=1}^{n}\xi_k$$

$$s.t.\quad y_k(w^\top x_k + b) \geq 1 - \xi_k \ (k = 1,2,\cdots,n)$$

$$\xi_k \geq 0 \ (k = 1,2,\cdots,n)$$

（2）解答：

SVM 的的学习策略就是间隔最大化，可形 式化为一个求解凸二次规划的问题，也等价于正则化的损失函数的最小化问题。 SVM 的的学习算法就是求解凸二次规划的最优化算法。

写出原问题的拉格朗日对偶函数

$$L(w,b,\xi,\alpha,u) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(y_i(wx_i+b)-1+\xi_i) - \sum_{i=1}^{n}u_i\xi_i$$

对w, b, ξ求偏导

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n}\alpha_i y_i \Phi(x_n)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow 0 = \sum_{i=1}^{n}\alpha_i y_i$$

$$\frac{\partial L}{\partial \xi} = 0 \Rightarrow C - \alpha_i - u_i = 0$$

将三式代入L中，得到

对上式求关于α的极大，得到：

$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i x_j) + \sum_{i=1}^{n}\alpha_i$$

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

$$\left.\begin{array}{l} C-\alpha_i-u_i=0 \\ \alpha_i \geq 0 \\ u_i \geq 0, i=1,2,\cdots,n \end{array}\right\} \Rightarrow 0 \leq \alpha_i \leq C$$

整理，得到对偶问题的最优化问题

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i x_j) - \sum_{i=1}^{n}\alpha_i$$

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

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

求得最优解α^*

$$w^* = \sum_{i=1}^{n} a_i^* y_i x_i$$

$$b^* = \frac{\max\limits_{i:y_i=-1} w^* x_i + \min\limits_{i:y_i=1} w^* x_i}{2}$$

实践中往往取支持向量的所有值取平均，作为 b*求得分离超平面

$$w^* x + b^* = 0$$

分类决策函数为

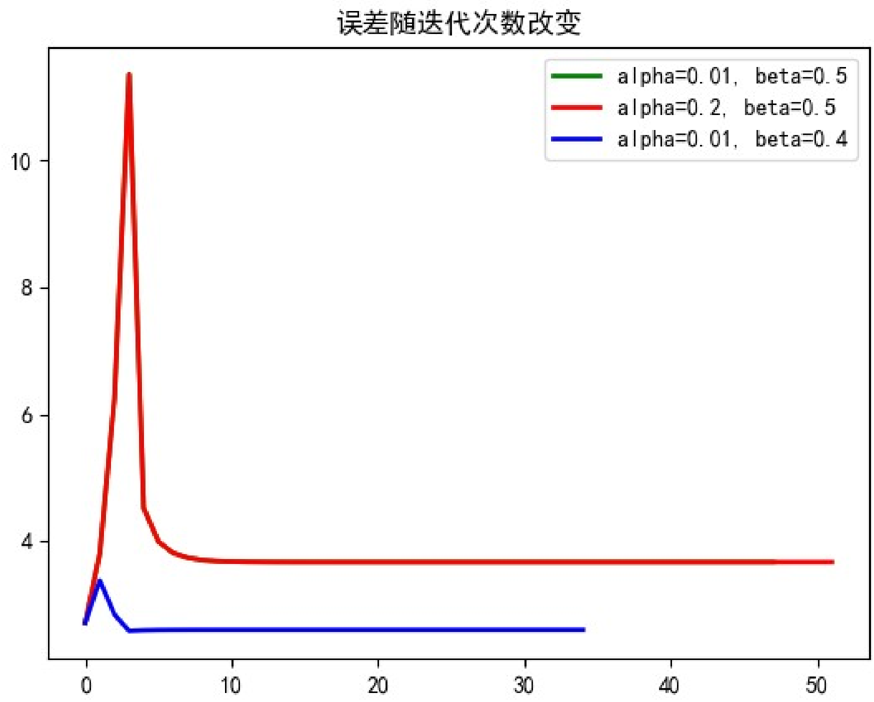
$$f(x) = sign(w^* x + b^*)$$

二、 非二次规划问题回溯直线搜索

（1）问题：

对于$\mathbb{R}^2$空间非二次规划问题 $\min f(x) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}$，分析回溯直线搜索采用不同的 α, β 值时，误差随迭代次数改变的情况。注：初始值相同。

（2）结果：



Beta 值对误差影响较大，beta 减小使误差减小

（3）工程文件

见附件 1

三、 等式约束熵极大化

（1）问题：

$$
\begin{array}{ll}
\text{minimize} & f(x)=\sum_{i=1}^{n} x_i \log x_i \\
\text{subject to} & Ax=b,
\end{array}
$$

其中 $\operatorname{dom} f=\mathbf{R}_{++}^{n}$，$A\in\mathbf{R}^{p\times n}$，$p<n$。（一些相关分析见习题 10.9。）

生成一个 $n=100$，$p=30$ 的问题实例，随机选择 A（验证其为满秩阵），随机选择一个正向量作为 $\hat{x}$（例如，其分量在区间 $[0,1]$ 上均匀分布），然后令 $b=A\hat{x}$。（于是，$\hat{x}$ 可行。）

采用以下方法计算该问题的解。

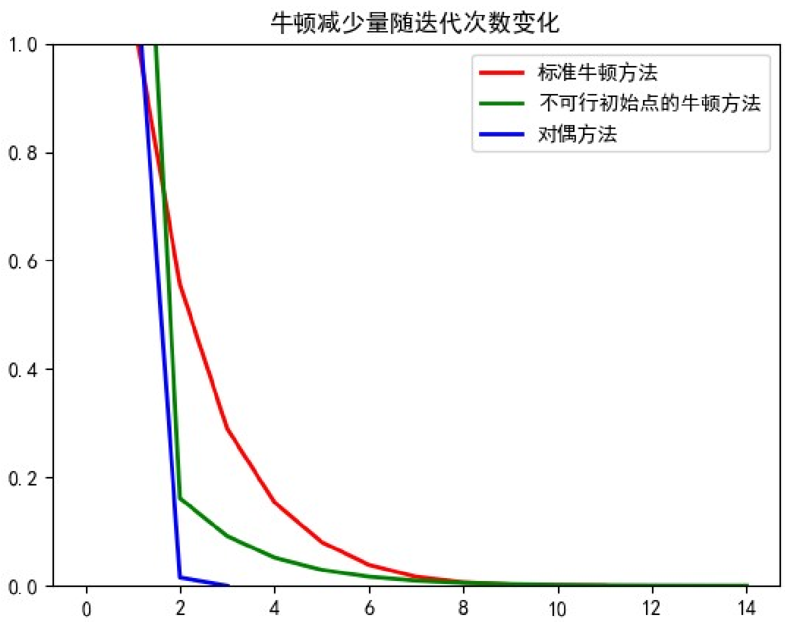
(a) **标准 Newton 方法**。可以选用初始点 $x^{(0)}=\hat{x}$。

(b) **不可行初始点 Newton 方法**。可以选用初始点 $x^{(0)}=\hat{x}$（和标准 Newton 方法比较），也可以选用初始点 $x^{(0)}=\mathbf{1}$。

(c) **对偶 Newton 方法**，即将标准 Newton 方法应用于对偶问题。

证实三种方法求得相同的最优点（和 Lagrange 乘子）。比较三种方法每步迭代的计算量，假设利用了相应的结构。（但在你的实现中不需要利用结构计算 Newton 步径。）

（2）结果：



从图像可以看出，对偶方法计算量最小，可行点的牛顿方法与不可行点的牛顿方法其次且相差较小。

（3）工程文件

见附件 2