总分 45 分。

一、填空题（16 分）

1. 闭凸集 $X \subset R^n$，任意一点 $y \in R^n$，则点 y 在集合 X 上的投影定义为：$Proj_X(y) =$ ________；$min_{x\in R^n}\frac{1}{2}||x-y||_2^2, s.t.\ Ax=b$，假设矩阵$A$列满秩，则其最优解$x^* =$________。
2. 在一维精确搜索方法中，二分法，黄金分割法和 Fibonacci 搜索都能将所搜索区间变窄，若以评估目标函数值的次数计算，评估次数越少则算法越有效，则这些搜索方法中效率最高的方法是：________。
3. 任意函数f的共轭函数定义为：$f^*(y) =$ ________。函数$f(x) = \frac{c}{2}x^2, c > 0$，其共轭函数表达式为：________。
4. 假设凸的扩展实值函数$f: R^n \to (-\infty, \infty]$,则函数$f$在点$x$处的次梯度$d$(表示列向量)应该满足的不等式为:$f(z) \geq$________,$\forall z \in R^n$。函数$f(x) = |x|$在$x = 0$处的次微分为：________。
5. $f(x) = \frac{1}{2}x^T Ax + b^T x + \mathrm{C}, A^T = A$，梯度$\nabla f =$________，Hessian 矩阵$\mathrm{H}(x)$=________。如果采用最速下降法求解该函数的极值，则其迭代公式为：________。如果采用牛顿法求解该函数的极值，则其迭代公式为：________。
6. 设目标函数$f(x)$具有连续的一阶偏导数，$x^{(k+1)}$由下列规则产生：

$$\begin{cases} f\left(x^{(k)} + \lambda_k p^{(k)}\right) = min_\lambda f\left(x^{(k)} + \lambda p^{(k)}\right) \\ x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)} \end{cases}$$

则有：$\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} p^{(k)} =$________。

7. $min\ f(x), s.t.\ g(x) \geq 0, h(x) = 0$，设 $x^* \in X, g_i(x)\left(i \in I, I\text{表示起作用约束集}\right), h_j(x), j = 1,2,\cdots,l$ 在 x^* 二次可微，$g_i(x), i \notin I$ 在 x^* 连续。约束规范$\{\nabla g_i(x^*), i \in I; \nabla h_j(x^*), j = 1,2,\cdots,l\}$线性无关。如果$x^*$是局部最优解，那么存在$u_i \geq 0 (i \in I), v \in R^l$,使得 KKT 条件式成立。则该问题的拉格朗日函数为：________。
8. $\min c^T x,\ s.t. \begin{cases} Ax = b, c \in R^n, A_{m\times n}, b \in R^m \\ x \geq 0 \end{cases}$的拉格朗日对偶问题为:________。
9. 二次终止性是指：________。为什么研究算法的二次终止性：________。

二、计算证明题（10:4 分，11：4 分，12 题：4 分，共计 12 分）

10. 请用大 M 法求解下列问题：

$$\max 2x_1 - x_2 + 2x_3$$

s.t. $\begin{pmatrix} -1 & -2 & 1 \\ 1 & -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$

11. 考虑如下非线性优化问题：

$$\min -7x_1 - 5x_2$$

$$\text{s.t.} \begin{cases} 2x_1^2 + x_2^2 + x_1 x_2 - 4 \le 0 \\ x_1^2 + x_2^2 - 2 \le 0 \\ -x_1 + \frac{1}{2} \le 0 \end{cases}$$

向量$\mathrm{x}_0 = (1,1)^T$,采用 KTT(Karush-Kuun-Tucker)定理来描述该点的 KTT 条件。验证该点是否是最优点。

12. 给定向量$w \in R^m$,令$w_{[|k|]}$表示向量按照绝对值从大到小排序后的第k个元素，即：$|w_{[|1|]}| \ge |w_{[|2|]}| \ge \cdots \ge |w_{[|m|]}|$.定义最大$L$范数为：

$$||w||_{[|L|]} \stackrel{\text{def}}{=} \sum_{k=1}^{L} |w_{[|k|]}, (L \in \{1,2,\cdots,m\})$$

请尝试证明，该范数是凸的（注意，常见的p范数中，$p \ge 1$都是凸的）。

三、问答题（13：4 分，14：4 分，15：2 分，共计 10 分）

13. 一般机器学习中的目标函数可以写为：

$$f^* \coloneqq \min_{x \in R^n} \{f(x) = \sum_{i=1}^{N} L(x^T u_i, v_i) + \lambda r(x)\},$$

这里值$\lambda \ge 0$，$L(\cdot,\cdot)$为代价函数，$r(\cdot)$为正则化函数。请列举出目标函数和正则化项分别采用不同的度量方式时（不同范数），所对应的典型问题，并结合自己的研究方向或者熟悉的研究方向来讨论其实际应用情况。

14. 请给出基于迭代的优化算法设计的基本出发点，给定一个复杂的非线性优化问题，一般如何去求解该问题。

15. 请总结无约束与有约束优化方法其解决思路有何异同。

四、论述题（7 分）

16. SVM 的基本原理：请分析一下采用超平面对样本点$\{(x_i, y_i)\}_{i=1}^N, x_i \in R^p, y_i\{-1,1\}$进行分类时，如何设计最大间隔分类器，又是如何利用拉格朗日函数来进行求解的？