没有系统学过数学优化，但是<u>机器学习</u>中又常用到这些工具和技巧，机器学习中最常见的优化当属凸优化了，这些可以参考Ng的教学资料：<u>http://cs229.stanford.edu/section/cs229-cvxopt.pdf</u>，从中我们可以大致了解到一些凸优化的概念，比如凸集，凸函数，凸优化问题，线性规划，二次规划，二次约束二次规划，半正定规划等，从而对凸优化问题有个初步的认识。以下是几个重要相关概念的笔记。

凸集的定义为：

Definition 2.1 *A set C is convex if, for any $x, y \in C$ and $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$,*

$$\theta x + (1 - \theta)y \in C.$$
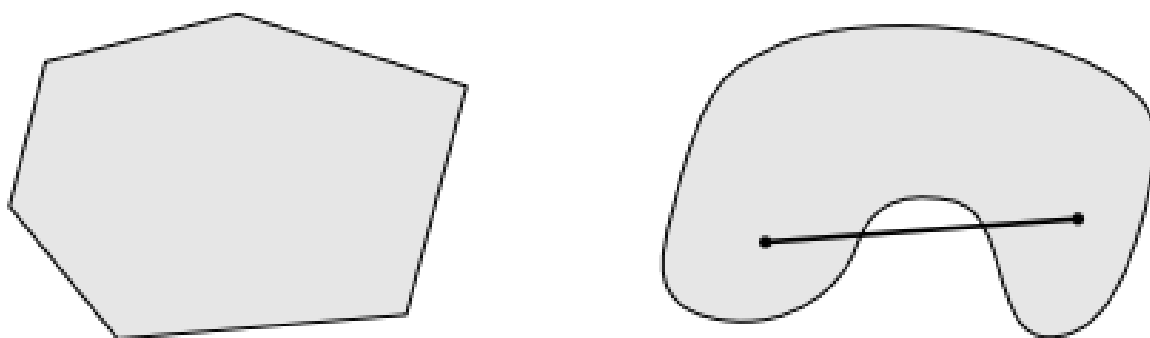
其几何意义表示为：如果集合C中任意2个元素连线上的点也在集合C中，则C为凸集。其示意图如下所示：

Figure 1: Examples of a convex set (a) and a non-convex set (b).

常见的凸集有：

n维实数空间；一些范数约束形式的集合；仿射子空间；凸集的交集；n维半正定矩阵集；这些都可以通过凸集的定义去证明。

凸函数的定义为：

Definition 3.1 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain (denoted $\mathcal{D}(f)$) is a convex set, and if, for all $x, y \in \mathcal{D}(f)$ and $\theta \in \mathbb{R}, 0 \leq \theta \leq 1$,*

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

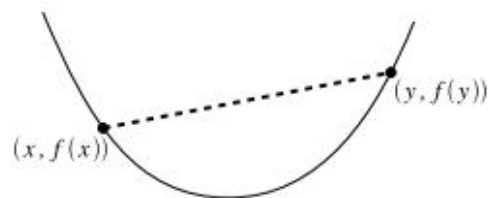其几何意义表示为函数任意两点连线上的值大于对应自变量处的函数值，示意图如下：

Figure 2: Graph of a convex function. By the definition of convex functions, the line connecting two points on the graph must lie above the function.

凸函数的一阶充要条件为：

$$f(y) \geq f(x) + \nabla_x f(x)^T (y - x).$$

其中要求f一阶可微。

二阶充要条件为：

$$\nabla_x^2 f(x) \succeq 0.$$

其中要求f二阶可微，表示二阶导数需大于0才是凸函数。

按照上面的两个定义，如果f(x)=x^2肯定是凸函数，而g(x) = -x^2是非凸函数。也就是说开口向下的函数是非凸函数，但是对于这种情况可以通过添加负号变成凸函数，从而求解。

常见的凸函数有：指数函数族；非负对数函数；仿射函数；二次函数；常见的范数函数；凸函数非负加权的和等。这些可以采用上面2个充要条件或者定义去证明。

凸优化问题（OPT）的定义为：

Armed with the definitions of convex functions and sets, we are now equipped to consider **convex optimization problems**. Formally, a convex optimization problem in an optimization problem of the form

$$\text{minimize} \quad f(x)$$
$$\text{subject to} \quad x \in C$$

where f is a convex function, C is a convex set, and x is the optimization variable. However, since this can be a little bit vague, we often write it as

$$\text{minimize} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \leq 0, \quad i = 1, \ldots, m$$
$$\qquad\qquad\quad h_i(x) = 0, \quad i = 1, \ldots, p$$

即要求目标函数是凸函数，变量所属集合是凸集合的优化问题。或者目标函数是凸函数，变量的约束函数是凸函数（不等式约束时），或者是仿射函数（等式约束时）。

对于凸优化问题来说，局部最优解就是全局最优解。

常见的凸优化问题包括：

线性规划（LP）：该问题是优化下面的式子：

- **Linear Programming.** We say that a convex optimization problem is a *linear program* (LP) if both the objective function f and inequality constraints g_i are affine functions. In other words, these problems have the form

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{subject to} \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

其中那个不常见的奇怪符号表示按元素小于等于，后面出现类似符号可以类似理解。

二次规划（QP）：该问题是优化下面的式子：

Quadratic Programming. We say that a convex optimization problem is a *quadratic program* (QP) if the inequality constraints g_i are still all affine, but if the objective function f is a convex quadratic function. In other words, these problems have the form,

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2} x^T P x + c^T x + d \\ \text{subject to} \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

二次约束的二次规划（QCQP）：该问题是优化下面的式子：

Quadratically Constrained Quadratic Programming. We say that a convex optimization problem is a *quadratically constrained quadratic program* (QCQP) if both the objective f and the inequality constraints g_i are convex quadratic functions,

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2} x^T P x + c^T x + d \\ \text{subject to} \quad & \tfrac{1}{2} x^T Q_i x + r_i^T x + s_i \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

半正定规划（SDP）：该问题是优化下面的式子：

Semidefinite Programming. This last example is more complex than the previous ones, so don't worry if it doesn't make much sense at first. However, semidefinite programming is becoming more prevalent in many areas of machine learning research, so you might encounter these at some point, and it is good to have an idea of what they are. We say that a convex optimization problem is a *semidefinite program* (SDP) if it is of the form

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}(CX) \\ \text{subject to} \quad & \operatorname{tr}(A_i X) = b_i, \quad i = 1, \ldots, p \\ & X \succeq 0 \end{aligned}$$

按照文章说SDP在机器学习领域应用很广，最近很流行，不过我好像没太接触到过。

参考资料：

http://cs229.stanford.edu/section/cs229-cvxopt.pdf