1. 有监督学习和无约束优化问题的优化方法分别有哪些？

第一类：直接解法。这种方法中只用到函数f（x），而不涉及其导数，

如坐标轮换法、鲍威尔法（Powell法）、随机搜索法、单纯形法等；

1. 坐标轮换法

坐标轮换法(univariate search technique)的思想极为朴素：从某个初始点x\_0出发，沿着各个坐标轴依次做一维搜索得x\_k，直到x\_k不再显著变化。

1. Powell法

Powell方法主要有三部分构成，即基本搜索、加速搜索和调整搜索方向。

2.1基本搜索：包括从基点出发沿着已知的 个线性无关的搜索方向进行一维搜索,确定一个新的基点;

2.2加速搜索：指这将第一轮起点和新确定的基点相连构成新的方向，进行一维搜索使得函数值下降；

2.3调整搜索：将这一轮方向组的第一个方向去掉,新的方向补充在最后构成新的方向组，进行下一轮迭代。

1. 随机搜索法
2. 单纯形法

第二类：间接解法。它要用到f（x）的导数，

如用到一阶导数的方法有梯度法、共扼梯度法和变尺度法；

用到二阶导数的方法以牛顿法为代表。间接解法也称为解析法。

1. 梯度下降法

梯度下降法（Gradient Descent Method）也叫做最速下降法（Steepest Descent Method），因为负梯度是函数局部下降最快的方向。

1. 共轭梯度法

由于每次都是沿着当前的负梯度方向逼近极小值，梯度下降法往往不能很快地收敛到极小值点。改进的方法是，在上一次梯度方向的共轭方向上进行迭代

1. 变尺度法
2. 牛顿法

牛顿方向和梯度方向最大的差别是考虑了Hessian矩阵

1. 拟牛顿法法

牛顿法的局限性基本源于∇^2f(x\_k)。在拟牛顿法中，我们不直接使用∇^2f(x\_k)，而是使用H\_k近似代替。

1. 随机梯度下降为什么会失效？

概念：随机梯度下降（sgd）

什么是随机梯度下降，怎么随机的呢？

其实就是在求梯度的时候，不再用所有的m个样本数据来计算，而是随机的选择一条数据来计算梯度！

随机梯度下降的好处是什么？缺点是什么？

在求梯度的时候快，迭代次数有可能更多，最终可能落不到全局最优解上

Mini-Batch GD是什么？

就是在求梯度的时候做了一个折中，不用所有的数据，而是随机选择一部分数据来求梯度！

机抽取一条数据来求解梯度，还随着迭代次数的增多，不断减小步长！learning\_rate

为什么要不断的调整步长？

就是为了让越接近最优解的时候，调整的幅度越小，避免来回震荡！

如果我们不人为的调小步长，会不会随着迭代的次数增多，调整的幅度自动减小？

调整的幅度取决于谁？却决于学习率和梯度，梯度事实上越接近最优解，梯度的绝对值越小