

Homework 2

钟赞 2016K8009915009

September 14, 2018

1 Answer Sheet

Theorem 1 寻找最佳 k -划分。

Proof 1 为了找出最佳的 k -划分，我们采用动态规划的方法。

设样本为 a_1, a_2, \dots, a_n 。

设 g 为目标函数， $g(a_i, a_j, t)$ 表示把 a_i 到 a_j 进行 t 划分的目标函数值。

考虑 x 为在 a_i 和 a_j 之间的样本，则有如下公式成立：

$$g(a_i, a_j, t) = \min_{a_i \leq x \leq a_j} (g(a_i, x, 1) + g(x, a_j, t-1)) \dots (*)$$

上述公式的正确性可由贪心算法决定，即希望小划分合起来组成的大划分的目标函数尽可能的小。
接下来我们先算小的划分，比如 1-划分：

$$g(a_i, a_j, 1) = \frac{-2B^2}{2A + \lambda} + \gamma$$

由对应的 A, B 的值，可以计算出每个 $g(a_i, a_j, 1)$ 的值。

再根据 (*) 式计算出所有的 2-划分 $g(a_i, a_j, 2)$ ，

$$g(a_i, a_j, 2) = \min_{a_i \leq x \leq a_j} (g(a_i, x, 1) + g(x, a_j, 1))$$

所有的 3-划分：

$$g(a_i, a_j, 3) = \min_{a_i \leq x \leq a_j} (g(a_i, x, 1) + g(x, a_j, 2)) \dots$$

依此类推可计到 $g(a_1, a_n, k)$ 。

由 (*) 式得到这个 k -划分必定是由很多个 1-划分合并起来的，我们根据计算的过程找出每次划分的点，这样就找到了最佳 k -划分。

下面是算法的伪代码：

- 1、对于 $t=1 \rightarrow n$ ，求出所有 $g(a_i, a_j, t)$ 的值
- 2、对于 $t=1$ ，求出所有 $g(a_i, a_j, 1)$ 的值
- 3、对于 $t=2 \rightarrow n$ ，所有 $g(a_i, a_j, t)$ 的值可由公式 (*) 求出：

$$g(a_i, a_j, t) = \min_{a_i \leq x \leq a_j} (g(a_i, x, 1) + g(x, a_j, t-1)) \dots (*)$$

并且记录每一个 t -划分的划分点。

- 3、以此类推，可以得到 k -划分的所有划分点以及它对应的目标函数的值。

该算法的复杂度为 $O(n^3)$ ，在多项式时间算法内。