机器学习第二次作业

周方全 | 202121081229　　邮件地址：2542154447@qq.com

I. 用线性回归处理非线性问题有什么办法，试举其中三个具体例子。

II. 什么是最小平方回归（Least Squares Regression），写出它的损失函数，试用从概率观点给出解释。请用它拟合下面三个数据点：(1,0.9)、(0,0.1)、(2,2)。

III. 什么是岭回归（Ridge Regression），写出它的损失函数，试用从概率观点给出解释，并说明它的优点。

IV. 什么是 LASSO 回归，写出它的损失函数，试从概率观点给出解释，并说明它的特点。

证明步骤：

首先，我们需要证明这个问题是 $\mathcal{NP}$ 问题。即，解决这个问题需要指数级的时间复杂度，验证结果则需要多项式级的时间复杂度。

其次，我们需要证明这个问题是 $\mathcal{NP}-Hard$ 问题。因为，$\mathcal{NP}-Complete$ 问题是 $\mathcal{NP}$ 问题和 $\mathcal{NP}-hard$ 问题的交集。

最后，我们需要找到一个已经存在的 $\mathcal{NP}-Complete$ 问题，如果能将它在多项式时间内归约到我们所求的问题上，说明我们的问题是 $\mathcal{NP}-Complete$。

V. Given a graph, a dominating set is a subset of vertices such that any vertex not in this set is adjacent to at least one vertex in this set The dominating set problem is to check whether a given graph has a dominating set of size at most k.

2.1 Prove that the dominating set problem is in $\mathcal{NP}$.

证明：

对于一个有 V 个顶点的图，他有 2^V 个子图。我们要解决支配集问题，要判断所有的 2^V 个子图才能找到这个问题的答案。而对于一个给结果而言，我们只需要判断他是否满足即可，我们首先判断支配集里面的顶点每个都是否相连，然后判断支配集外面的顶点是否和支配集里面的顶点相连。而这个判断过程在多项式时间复杂度内就可以完成。

综上，解决这个问题需要指数级的时间复杂度，而验证这个问题则需要多项式时间复杂度，所以他是 $\mathcal{NP}$ 问题。

2.2 Prove that the dominating set problem is NP hard.

证明：

已知：顶点覆盖问题 $Vertex-coverProblem$ 是 NP-完全问题。

对给定的无向图 G，作如下处理：对于图 G 的任意一边 (u,v)，添加一个点 w，使得该边的两个顶点 u□v 分别与 w 相邻，得到新的一个无向图 G'。

1. 若原无向图 G 存在一个顶点覆盖 S，且 S 满足 $|S| <= k$，则 S 也可以作为图 G' 的一个满足条件的支配集。原因如下：

假设，顶点集 S 是图 G 的一个点覆盖，而却不是图 G' 的一个支配集。那么会存在一个点 $w \in V(G')$，且 $w \notin S$ 但是 w 不和 S 任意一个点有相连边。因为在我们构造的 G' 中，顶点 u, v, w 是彼此连通的，因为 w 不在 S 中，那么顶点 u, v 一定在 S 中。但是边 (u,v) 的两个端点都在 S 中。显然和假设 S 是一个点覆盖相矛盾。那么，S 不是图 G' 的一个支配集是错误的。所以，无向图 G 存在一个顶点覆盖 S，且 S 满足 $|S| <= b$，则 S 也可以作为图 G' 的一个满足条件的支配集。

2. 若无向图 G' 存在一个支配集 D，且满足 $|D| \leq k$，则图 G 存在满足条件的顶点覆盖 S

对于图 G' 中的任意一条边 (u,v)，以及它的扩展顶点 w。构成一个三元式 $\{u,v,w\}$。

若 $w \notin D$，那么为了满足 D 是一个支配集，u□v 中的只能有一个且一定有一个在 D 中，那么图 G' 的支配集 D 就是图 G 的顶点覆盖 S；

若，$w \in D$，为了满足 D 是在一个支配集，u□v 两个顶点不能在 D 中。则将 D 中的顶点 w 替换成顶点 u 或 v，可以得到原图 G 的一个顶点覆盖 S。

VI. Prove that: if we can check whether a graph has a clique (a complete graph) of size k in polynomial time then we can also find a clique of size k in polynomial time.

证明：

由题意得：设函数 $fun(G,k)$，当图 G 中有一个存在一个大小为 k 的团的时候返回 $True$，否则返回 $False$。

Algorithm 1: 找到大小为 K 的团

Input: 一张图 $G(V,E)$, 团的大小 k

Output: 团大小为 k 的顶点集合

if $fun(G,k) \neq True$ **then**
    不存在这样的团;
    return None;
for 顶点 v *in* V **do**
    // 删除顶点 v 和其的相邻边
    $V' = V - \{v\}$;
    $E' = E - \{(u,v) : u \in V\}$;
    if $fun(G,k) == True$ **then**
        $V = V'$;
        $E = E'$;
    if $|V| == k$ **then**
        break;
return V ;

通过上述算法，我们可以得到大小为 K 的团。因为 $fun(G,k)$ 是多项式时间的算法。那么得到大小为 K 的团的算法的复杂度也是多项式时间的。

VII. A graph is called a 2 plex if each vertex in the graph is not adjacent to at most one other verte x. Prove that it is $\mathcal{NPC}$ to check whether an input graph has a sub graph of at least k vertices that is a 2 plex.

VIII. In the multiway cut problem, we are given a undirected graph $G = (V, E)$ and some special vertices in V (called terminals). The problem asks us to delete the minimum number of edges from the graph such that no pair of terminals is connected. Please give a 2-approximation algorithm for this problem.