

Selection Problems and the Greedy Algorithm

A *selection problem* is the problem of choosing a subset of items satisfying some given property from a finite set. If each element has a positive weight, the *maximum weight selection problem* is to find such a subset for which the sum of the weights is maximized. Specifically we consider problems with the following structure.

P1. S is a finite non-empty base set of items. Each subset $B \subseteq S$ is either *admissible* or *inadmissible*.

P2. (Hereditary Property). If B is an admissible subset and $A \subseteq B$ then A must also be admissible.

Many texts use the terms *independent* and *dependent* for admissible and inadmissible, respectively. Here are a few examples.

Example 1: Sushi problem. $S = \{1, 2, \dots, n\}$ is a set of n pieces of sushi, and a_i is the cost of piece i . You have b yen to spend. An admissible subset of sushi is a subset of S that costs at most b yen.

Example 2: Matchings in a graph. $G = (V, E)$ is a graph. $S = E$ is the edge set of the graph. A subset $B \subseteq S$ is admissible if no two edges in B share a common vertex (such a set of edges is called a matching).

Example 3: Committee selection. S is a set of n people and k is a positive integer. An admissible subset (committee) is a subset $B \subseteq S$ with $|B| \leq k$.

Example 4: Matrix columns. Let A be an m by n matrix. S is the set of column indices. A subset of columns indexed by $B \subseteq S$ is admissible if the columns it indexes are linearly independent.

Example 5: Travelling salesman tours. $G = (V, E)$ is a graph with n vertices. $S = E$ is the edge set of the graph. A subset B of edges is admissible if can be extended to give a cycle of length n (travelling salesman tour) in the graph. Ie., if B is not already a cycle of length n , we can add edges from $S \setminus B$ to B to make it one.

Example 6: Forests in a graph. $G = (V, E)$ is a graph with n vertices. $S = E$ is the edge set of the graph. A subset B of edges is admissible if it does not contain any cycle (such sets are called forests.)

Exercise: For each of these examples verify that it satisfies the hereditary property.

An admissible set B is called *maximal* if for each $i \in S \setminus B$, $B \cup \{i\}$ is inadmissible. In general we will be interested in selecting maximal admissible subsets. We get a *weighted selection problem* by assigning a weight $w_i > 0$ to each $i \in S$. The *weighted selection problem* is to find a maximal admissible subset with maximum (or minimum) total weight. Of course, since the weights are positive a maximum weight admissible subset is always maximal. However to make the minimum weight selection problems non-trivial, we need to specify that the subset required is maximal. We can extend each example by defining an appropriate weight for each element of S .

Example 1: Each piece of sushi i has a taste score $w_i > 0$. Find the plate of sushi costing less than b yen with maximum taste score. This is an example of the *knapsack problem*

Example 2: Let each edge $e \in S$ have a desirability value $w_e > 0$. Find the matching of maximum total weight. If G is a complete bipartite graph $K_{n,n}$, then this becomes the *assignment problem*.

Example 3: Let $w_i > 0$ be the number of votes obtained by person i in an election. Find the committee with maximum total number of votes.

Example 4: Let each column i have weight $w_i > 0$. Find maximum weight subset of linearly independent columns.

Example 5: Let each edge e have length w_e . Find the travelling salesman tour of minimum total length. This is called the *travelling salesman problem*.

Example 6: Let each edge e have length w_e . Find the maximal forest of minimum total length. If the graph G is connected, the maximal forests are spanning trees in the graph. The problem is then called the *minimum weight spanning tree problem*.

Some of the above problems are easier to solve than others. We will first look at perhaps the simplest possible selection algorithm, called the *greedy selection algorithm*.

```

T = ∅
while S ≠ ∅
    { choose maximum (or minimum) weight element x in S
      S ← S \ { x}
      if T ∪ {x} is admissible T ← T ∪ {x}
    }
return T;

```

Exercise: Find instances where the greedy selection algorithm fails to find a maximum (or minimum) weight maximal admissible set for Examples 1,2, and 5.

You should however find that greedy selection seems to work for Examples 3,4, and 6. The question is why? It always returns a maximal admissible set, but an additional property is required to guarantee that this set has maximum (or minimum) weight.

P3. (Exchange property) Let A and B be admissible subsets with $|A| < |B|$. Then there is always some $x \in B \setminus A$ such that $A \cup \{x\}$ is admissible.

Exercise: Show that Examples 3,4,6 satisfy P3 but that Examples 1,2,5 do not.

Property P3 is strong: for one thing it implies that all maximal admissible sets have the same cardinality (why?). Together with P1 and P2 it is enough to guarantee that the greedy selection algorithm always gives an optimum solution, as we now show.

Theorem The greedy selection algorithm gives an optimum solution for any weighted selection problem with base set S satisfying properties P1, P2 and P3.

Proof: We prove the theorem for the maximization problem, the minimization problem being similar. The proof is by contradiction: assume we have a selection problem satisfying properties P1, P2 and P3 for which a greedy solution $T \subseteq S$ has smaller weight than an optimum $O \subseteq S$. First observe that T and O are both maximal admissible sets, and so by P3 they must have the same cardinality, say r . Let $O = \{o_1, o_2, \dots, o_r\}$ and $T = \{t_1, t_2, \dots, t_r\}$ be labelled such that $w_{o_1} \geq w_{o_2} \geq \dots \geq w_{o_r}$ and $w_{t_1} \geq w_{t_2} \geq \dots \geq w_{t_r}$. Since greedy selection failed we must have

$$\sum_{i=1}^r w_{o_i} > \sum_{i=1}^r w_{t_i}.$$

This implies that there must be a smallest integer k such that $w_{o_k} > w_{t_k}$. Consider the sets

$$B = \{o_1, o_2, \dots, o_k\} \subseteq O, \quad A = \{t_1, t_2, \dots, t_{k-1}\} \subseteq T.$$

Both of these sets are admissible (P2) and so since $|B| > |A|$ there is some element $o_i \in B \setminus A$ such that $A \cup \{o_i\}$ is admissible (P3). Now $w_{o_i} \geq w_{o_k} > w_{t_k}$ so the element o_i must have already been considered by greedy selection. Since it is admissible to A it is admissible to any subset of A (P2) so it should have been chosen by greedy at that time. This contradiction proves the theorem. \square