CS/ECE 473 Fall 2020

Homework 9 Problem 1

Ryan Prendergast (ryanp4)

Noah Watson (nwatson3)

Lawson Probasco (lawsonp2)

(a) Write an integer linear programming formulation for k-median.

Solution: We define a linear programming with two types of variables according to the problem description. The first is $y_q, q \in P$. $y_q \in \{0, 1\}$ takes the value one if the point q is one of the k centers, and it takes the value zero otherwise. There are n such variables.

The second type of variable is $x_{p,q}, p, q \in P$. $x_{p,q} \in \{0, 1\}$ takes the value one if the point p is assigned to the point q. It takes the value zero otherwise. There are n^2 such variables

Our objective is to minimize the sum of the distances from each point to the point which it is assigned to. We must in the constraints ensure that there are exactly k centers, each point p is assigned to exactly 1 other point, and that if p is assigned to q, q is a center (in other words, if $x_{p,q} = 1$, then $y_q = 1$ as well).

$$min \sum_{p \in P} \sum_{q \in P} d(p,q) * x_{p,q} \qquad \text{Min the sum distances to cluster centers}$$

$$\text{s.t.}$$

$$\sum_{p \in P} y_p = k \qquad\qquad \text{There must be exactly k centers}$$

$$\sum_{q \in P} x_{p,q} = 1, \forall p \in P \qquad\qquad \text{Each point belongs to one center}$$

$$x_{p,q} \le y_q, \forall p, q \in P \qquad\qquad \text{Each point can only belong to a center}$$

$$x_{p,q}, y_q \in \{0, 1\} \qquad\qquad \text{Our variables take values 0 or 1}$$

The optimal solution to k-median is equivalent to the above linear program. Our constraints ensure that, as with k-median, there are k cluster centers decided by those that minimize the sum point distances to the cluster they are assigned to.

We can also relax the constraint that the x variables must be integers, making the simple change to our objective function and our constraints to replace $x_{p,q}$ with $\mathbb{1}[x_{p,q} > 0]$. Thus, any real positive real number is equivalent, instead of us checking with the value 1. ∎

(b) Write the dual of the LP relaxation.

Solution: We define the variables of the dual $a_{p,q}$ and b_q which correspond to x and y,

respectively.

$$max \sum_{p \in P} \sum_{q \in P} d(p,q) * a_{p,q}$$

$$s.t.$$

$$\sum_{q \in P} b_q * a_{p,q} \leq d(p,q) \forall p \in P$$

$$a_{p,q} \geq b_q, \forall p, q \in P$$

∎