Lista 4 zadanie 8

Kwiecień 2020

1 Treść zadania

Doktor Zseimel postanowił zaciągnąć się do pracy na platformie wiertniczej. Niestety został zakwalifikowany do kategorii osób o statusie “overqualified” i nie dostał wymarzonej pracy. Zaproponowano mu natomiast zostanie konsultantem strategicznym koncernu naftowego, który planuje budowę dużego rurociągu przebiegającego z zachodu na wschód przez pola naftowe, na których znajduje się n wież wiertniczych. Do każdej wieży ma dochodzić odnoga głównego rurociągu po najkrótszej możliwej ˙ drodze (albo na północ, albo na południe). Jakiego algorytmu powinien użyć doktor dla zadanych współrzędnych $(x_i, y_i)i = 1...n$ wież aby wyznaczyć optymalne położenie głównego rurociągu (czyli takie, dla którego suma długości odnóg jest minimalna)? Wykaż że można takie położenie wyznaczyć w czasie liniowym.
Osobiście założyłem że ”z zachodu na wschód” oznacza że rurociągu biegnie równolegle do równoleżnika ziemi.

2 Pierwsze wnioski

Jest to ewidentnie zadanie optymalizacyjne. Trzeba znaleźć taką prostą y=c (a = 0 bo jest to prosta równoległą do osi OX). Znamy wzór na odległość punktu(x, y) od prostej(ax + by + c = 0):

$$d = \frac{|ax + by + c|}{\sqrt{a^2 + b^2}}$$

W naszym przypadku to a = 0, b = -1, i chcemy ustalić c, a zamiast 1 punktu mamy ich n. Tak więc funkcja którą będziemy optymalizować wygląda:

$$d(c) = \sum_{i=1}^{n} |y_i - c|$$

Musimy zatem znaleźć minima lokalne naszej funkcji d. Z definicji wiemy że

$$\text{m jest minimum lokalnym} \iff (\exists\delta > 0)(\forall|\epsilon| < \delta)(d(m + \epsilon) \geqslant d(m))$$

Niech α oznacza liczbę punktów nad prostą($y = c$) i β liczbę punktów pod prostą i γ liczbę na prostej. Weźmy takie $\Delta c > 0$, że dla prostej $y = c + \Delta c$ znajduje się α punktów nad i $\beta + \gamma$ punktów pod, a dla prostej $y = c - \Delta c$ znajduje się $\alpha + \gamma$ punktów nad i β punktów pod. Wiemy zatem że:

$$d(c + \Delta c) = d(c) + \Delta c(\beta - \alpha + \gamma)$$

$$d(c - \Delta c) = d(c) + \Delta c(\alpha - \beta + \gamma)$$

Tak jak napisałem wcześniej szukamy takiego c że:

$$(\exists \Delta c)(d(c - \Delta c) \geqslant d(c) \leqslant d(c + \Delta c))$$

Rozpisując to używając α, β, γ:

$$d(c - \Delta c) \geqslant d(c) \wedge d(c + \Delta c) \geqslant d(c)$$

$$d(c) + \Delta c(\alpha - \beta + \gamma) \geqslant d(c) \wedge \Delta c(\beta - \alpha + \gamma) \geqslant d(c) + d(c)$$

$$\Delta c(\alpha - \beta + \gamma) \geqslant 0 \wedge \Delta c(\beta - \alpha + \gamma) \geqslant 0$$

$$\alpha - \beta + \gamma \geqslant 0 \wedge \beta - \alpha + \gamma \geqslant 0$$

$$\gamma + \alpha \geqslant \beta \wedge \gamma + \beta \geqslant \alpha$$

Widzimy zatem że musimy znaleźć takie c, liczba punktów nad prostą $y = c$ była nie większa niż liczba pozostałych punktów, tak samo liczba punktów pod prostą nie była większa od liczby pozostałych punktów.

3 Mediana

Pokażmy że mediana zbioru X nazwijmy ją $\tilde{x}$ ma takie właściwości że, liczba elementów mniejszych od niej jest nie większa niż liczba elementów nie mniejszych, tak samo liczba elementów większych od niej jest nie większa niż liczba elementów nie większych.
Załóżmy że $|\{x \in X : x < \tilde{x}\}| > |\{x \in X : x \geqslant \tilde{x}\}|$. Wynika z tego że mediana jest albo elementem zbioru $\{x \in X : x < \tilde{x}\}$, albo średnią dwóch elementów z tego zbioru. Ale ten zbiór zawiera tylko elementy mniejsze od od mediany, a więc sprzeczność.
Identyczny dowód dla elementów większych.

4 Minimum globalne

Wiemy dodatkowo że ta funkcja nie ma właściwych maksimów lokalnych. Ponieważ maksima lokalne muszą spełniać:

$$\alpha - \beta + \gamma \leqslant 0 \wedge \beta - \alpha + \gamma \leqslant 0$$

$$\gamma \leqslant -(\alpha - \beta) \wedge \gamma \leqslant (\alpha - \beta)$$

$$\gamma \leqslant -|\alpha - \beta|$$

A wszystkie punkty spełniające to są również minimami lokalnymi:

$$(\gamma \leqslant -|\alpha - \beta|) \iff (\gamma = 0 \wedge |\alpha - \beta| = 0) \Rightarrow (\gamma + \alpha \geqslant \beta \wedge \gamma + \beta \geqslant \alpha)$$

A co za tym idzie każde minimum lokalne jest również minimum globalnym.

5 Rozwiązanie

Widzimy zatem że mediana zbioru $\{y_1, y_2, ..., y_n\}$ ma właściwości spełniające warunki bycia minimum funkcji $d(c)$. Ponieważ jesteśmy w stanie ją określić w czasie $O(n)$ używając algorytmu select (poznanym na wykładzie) to jesteśmy również w stanie rozwiązać ten problem w liniowym czasie.