CS6100 – Topics in Design and Analysis of Algorithms

Beyond Worst Case Analysis

Instructor:

Raghavendra Rao B V (BSB 354, bvrr@cse.iitm.ac.in, rkunjathaya@gmail.com)

Lecture 2

2 Opt algorithm for TSP

TSP Given n-cities obtain a min. wt Hamiltonian cycle.

$G = (V, E)$ with $w: E \to \mathbb{R}^+$

2-opt $\longrightarrow$ k-opt

$e = \{x_1, x_2\}$

$f = \{y_1, y_2\}$

$x_1 \ x_2 \ \ldots \ y_1 \ y_2$

1. Let T be a tour
2. For every pair of edge e, f in T see if the corresp. 2-opt step reduces the cost of the tour. If yes then update the tour.
3. repeat (2) until the tour stabilizes
4. Output T.

either T will not be optimal or (2) will be run super-poly time

Local improvement heuristics.

Known

① There are input instances when 2-opt runs for $2^{\Omega(n)}$ many iterations

c

α-optimal

$\frac{c}{\alpha} \geq 1$

Quality of o/p

Any general TSP, there is no $n^{O(1)}$ factor approximation algorithm unless P = NP

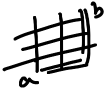
This extends to 2-opt (Chandra et al)

for ETSP $\frac{c}{\alpha} \leq O(\log n)$

$\exists$ instance s.t. $\frac{c}{\alpha} \geq \frac{\log n}{\log \log n}$

Variants of TSP

Euclidean TSP   metric TSP

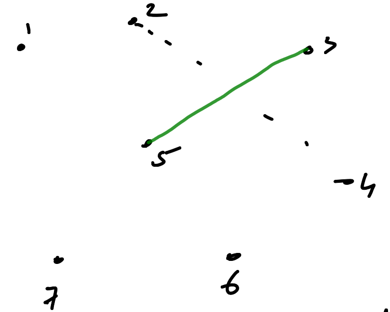


Assume that the cities are points in a plaine

distance: std Euclidean distance.

$p = (x_1, y_1)$

$q = (x_2, y_2)$



$d(p,q)$ – l_2-norm

$= \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$

Properties:- $d(p,p) = 0$

$d(p,q) \leq d(p,w) + d(w,q)$ $\longrightarrow$ Δ^{le} inequality

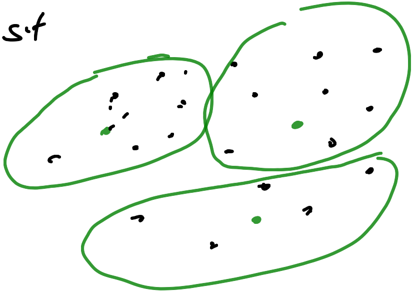
k-means clustering

i/p: set of n-points $p_1, \dots, p_n$ & $k > 0$

o/p: A partition $C_1, \dots, C_k$ of the points along with centers $\beta_1, \dots, \beta_k$ s.t

$$\sum_{i=1}^{k} \sum_{x \in C_i} d(x, \beta_i)$$

is minimized



$(1, 2)$ $(3, 4)$ $(1, 5)$ — C_1

$\mu = \left(5/3, 11/3\right)$

K-means method

initialize with arbitrary $C_1, \ldots, C_k$

$\text{mean}(C_i) = \frac{1}{|C_i|} \sum_{x \in C_i} x$

① set $\mu_i = \text{mean}(C_i)$

② re-adjust the cluster:
$\forall p$, p goes to its nearest center.

③ repeat ① & ② until no improvement

— Status of K-means is similar to that of 2-opt

Average Case analysis

A algo, t - measure.

F_n: all inputs of length n.

Performance of algo

$$avg_A(n) = \frac{1}{|F_n|}\left(\sum_{x \in F_n} t(x)\right)$$

↳ work if F_n is finite

$$\max_{x \in F_n} t(x)$$

Avg case complexity of QS $= O(n \log n)$

When F_n is not finite

$$avg_A(n) = \mathbb{E}_{x \sim F_n}[t(x)]$$

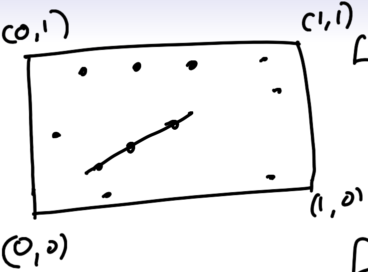
- Probabilistic analysis

ETSP - or k-means in Euclidean domain

n-points

$p_1 \quad p_2 \quad p_3$

$c\sqrt{n}$



$[0,1]^2$

$[0,1]$

$[0,1]^2$

Beyond worst case

Limitations of average case analysis

Limitations of average case analysis

Smoothed Analysis