CS4786/5786: Machine Learning for Data Science, Spring 2015
3/12/2015: Lecture 13 handout: Kleinberg's (2003) impossibility theorem for clustering

---

Announcement Due to the upcoming drop deadline, we plan to release A1 grades on CMS this afternoon or tonight, regardless of whether all the grades have been entered. If you want to know when *your* grade is entered, set your notifications on CMS to be sent an email when "...one of your grades is changed" (and perhaps also "...grades for an assignment are released")

I. Clicker questions (optimal clustering when k can vary) Let k^* be the number of clusters in the optimal clustering for the given question. Assume $n >> 2$, where n is the number of data points, as usual. What is k^* for ...

Question	*Answer choices*
I **(a)** ... the k-means optimization function?	(A) 1 (B) 2 (C) $n-1$ (D) n (E) I don't know
I **(b)** ... normalized cut?	
I **(c)** ... the single-link optimization function (disallowing $k^* = 1$)?	

II. Potential stopping conditions for single-link Which is the most sensible condition for adapting single-link clustering to produce different numbers of clusters?

(A) (distance-α) Let α be a fixed positive number. Stop adding edges (i.e., merging clusters) when there are no more cross-cluster pairs of points with distance $\geq \alpha$.

(B) (scale-α) Same as above, except change the last part to "when there are no more cross-cluster pairs of points with distance $\leq \alpha \times \max_{t,s} ||\mathbf{x}_t - \mathbf{x}_s||_2^2$."

(C) Exactly one is sensible, but both would be good if you changed the direction of one of the inequalities

(D) Neither are good, but both would be if you changed the direction of both the inequalities

(E) What's single-link clustering?

Let $X = \{1, 2, \ldots\ldots, n\}$.
Clustering functions f have:

input: a distance function $d : X \times X \to \Re^{\geq 0}$
output: a partition of X

III. Richness property $\mathrm{Range}(f)$ = set of all partitions of X.

IV. Scale-invariance property For any distance function d and any $\alpha > 0$, let $d_\alpha(t, s) \stackrel{def}{=} \alpha \cdot d(t, s)$. Then $f(d) = f(d_\alpha)$.

V. Consistency property Let d be any distance function.
Let $P = f(d)$ be the partition that is output.
Let d' be any distance function where $(t, s) \sim P$ implies $d'(t, s) \leq d(t, s)$,
and
where $(t, s) \not\sim P$ implies $d'(t, s) \geq d(t, s)$,
Then, $f(d') = f(d)$.

VI. Properties of known clusterings Which do various clustering algorithms satisfy, of
(A) just richness
(B) just scale invariance
(C) just consistency
(D) exactly 2
(E) all 3

VI **(a)** fixed-k single link

VI **(b)** distance-α single link

VI **(c)** fixed-k-means

Theorem 3.1 Let f satisfy scale-invariance and consistency. Then Range(f) is an *anti-chain*.

Theorem 3.2 For every anti-chain A, there is an f such that Range(f) $= A$ and f satisfies scale-invariance and consistency.