Last time, we started exploring the idea of clustering, a form of dim. reduction.

- Looked at functions for describing how good a clustering is, and
  explored relationships between them.

- But we didn't talk about how you get a good clustering (i.e.,
  didn't describe any algorithms).

There's a good reason for this.

Note that we're talking about a huge search space.

<see "Stirling numbers of the second kind";
  bounded below by \( \frac{1}{2} (k^2 + k + 2) / (k-1) \), according to Wikipedia>

Hatges webinar: Mckearn '92: n=21, k=9, 1.23x10^4
... although that doesn't mean that clustering is hard.
<sorting also has a huge search space>

Today, two alg.

<follow textbook (2nd page)

(i) seems to be C6(a), just to compute the criterion

- Note that (2) requires being able to compute centroids - if your data can't gain as
  points in R^k, it might not be able to do this. Example: only given point-to-point
  distances.

Visualization in 3D: www.mathalinos.com/blog/visualizing-k-means-clustering

Start with packed circles

- Explain why closest, @ = centroid, colors = closest pt whose closest centroid
  is that color's.

(need someone to "put finger on r_j\(^{-1}\))

(in class, accidentally picked an r_j that became empty... oops.)

\(\ast\) Specifically: can say k-means opt is inspired by k, but it is actually:

- In (8), rewrite: \( \sum_{x \in C_j} ||x - r_j\|^2_2 \), so r_j may not always be centroid;

- Now it's of (8). Rewrite:

\[
\sum_{x \in C_j} ||x - r_j\|^2_2 = \sum_{x \in C_j} ||x - x_j\|^2_2 + n_j ||r_j - x_j\|^2_2
\]

Last time: note that it seems like this is a violation of \( \Delta \Phi \):

<click results>
we resolve the paradox by noting that this is the Euclidean distance squared.

Let $a, b, c$ be 2 triangle arm-lengths, $c$ the length of "connector".

$$a^2 + b^2 = c^2$$

For right angled, $c = \sqrt{a^2 + b^2}$.

So, now that we at least don't think that (8) is false a priori, why don't we go ahead and prove it?

$$\sum_{x \in C_j} \| x - r_j \|^2$$

need to introduce $r_j$. Let's break-force it.

$$= \left( \sum_{x \in C_j} \| x - r_j \|^2 \right) - \langle \sum_{x \in C_j} (x - r_j), (r_j - g) \rangle + \sum_{x \in C_j} \| r_j - g \|^2$$

$$= \left( \sum_{x \in C_j} \| x - r_j \|^2 \right) - \langle \sum_{x \in C_j} (x - r_j), (r_j - g) \rangle + \sum_{x \in C_j} \| x - r_j \|^2$$

$$= \left( \sum_{x \in C_j} \| x - r_j \|^2 \right) - \langle \sum_{x \in C_j} (x - r_j), (r_j - g) \rangle + \sum_{x \in C_j} \| r_j - g \|^2$$

Second term is length.

Making it a good choice for $r_j$.  

- Corollary: the $g$ that minimizes either LHS / RHS is the centroid.
and again, we assert what proof that this lemma yields
that (1) => (2). See Hopcroft, Kannen (link on wiki) for pf.

So: Considering RHS:

\[ \text{(K-means) style 1 reduces distance to the representative:} \]

\[ \text{style 2 minimizes w.r.t. cluster minimizes per cluster terms.} \]

\[ \text{[are we playing first and last or first of all?] ?} \]

We were choosing the best \( f \)'s for the \( C_j \) fixed \( C_j \)s.

\[ \text{convergence gain by feasible improvements in LHS having to be in discrete} \]

\[ \text{increments (if we assume no cycling between clusterings).} \]

\[ \text{(this alternating minimization)} \]

\[ \text{does it optimize? no (NP-hard problem).} \]

\[ \text{partitioning problems are "pretty much always" NP-complete.} \]

--- except when they aren't: lead-in to single-link clustering: an optimal alg.

\[ \text{(for a certain clustering criterion).} \]

\[ \text{note: "best friend" criterion turns out to not be the single-link criterion.} \]

\[ \text{(probable) - possible bug exercise! Interesting to understand these subtleties.} \]

\[ \text{fig} \]

instead: take \( \delta \) inter-cluster spacing (front of handout).

\[ \max \text{ the minimum of } d(C_j, C_j') = \min_{x \in C_j, x' \in C_j'} \| x - x' \|^2 \]

\[ \text{note that not all pairwise intra-cluster distances are small.} \]

\[ \text{2-cluster} \]

\[ \text{idea of single-linkage: "knock out" the minimum spacing} \]

\[ \Rightarrow \text{the max increases} \]

\[ \text{but what about optimality?} \]