10-806 Foundations of Machine Learning and Data Science

Lecturer: Avrim Blum 11/11/15

1st-half of class: The Johnson-Lindenstrauss Lemma

The Johnson-Lindenstrauss Lemma:

Given m points in R^n, if project randomly to R^k, for $k = O(\frac{1}{\epsilon^2}\log\frac{m}{\delta})$ then whp all pairwise distances preserved up to $1 \pm \varepsilon$ factor (after scaling by $\sqrt{n/k}$).

So, if we just care about apx distances, can convert high-dimensional data to moderate-dimensional data.

The "$\log m$" is just from union bound over the $\frac{m(m-1)}{2}$ pairs, so can replace with $k = O(\frac{1}{\epsilon^2}\log\frac{1}{\delta})$ if OK with "most pairs".

Say points are $p_1, p_2, \ldots, p_m$. Will ignore δ from now on.

JL Lemma, cont

Given m points in R^n, if project randomly to R^k, for $k = O(\varepsilon^{-2} \log m)$, then whp all pairwise distances preserved up to $1\pm\varepsilon$ (after scaling).

Proof easiest for slightly different projection:

- Pick k vectors $u_1, \ldots, u_k$ iid from n-diml Gaussian.
- Map $p \to (p \cdot u_1, \ldots, p \cdot u_k)$.
- What happens to $v_{ij} = p_i - p_j$?
  - Becomes $(v_{ij} \cdot u_1, \ldots, v_{ij} \cdot u_k)$
  - Each component iid from 1-diml gaussian, scaled by $|v_{ij}|$.
  - What happens to $\|\cdot\|^2$? For concentration on sum of squares, plug in version of Hoeffding for RVs that are squares of Gaussians.
- So, whp all lengths apx preserved, and in fact not hard to see that whp all <u>angles</u> are apx preserved too.

Random projection and margins

Natural connection:

- Suppose we have a set S of points in the unit ball in R^n, separable by margin γ.
- JL lemma says if project to random k-dimensional space for $k=O(\gamma^{-2} \log |S|)$, whp still separable (by margin $\gamma/2$).
  - Think of projecting points and target vector w.
  - Angles between p_i and w change by at most $\pm\gamma/2$.
- Could have picked projection before sampling data.
- So, it's really just a k-dimensional problem after all. Do all your learning in this k-diml space.

So, large margin implies in a sense it's really a lower-dimensional problem