COMS 4772 Fall 2015

Lecture 14

Agenda

- Sketch-and-solve algorithm for low rank approximation
- k-center and k-means clustering

Recall: approximate least squares

- Least squares problem:

$$\text{Given } \boldsymbol{A} \in \mathbb{R}^{n\times d} \text{ and } \boldsymbol{b} \in \mathbb{R}^{n} \ (n \gg d),$$

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \,.$$

- Instead of solving $n\times d$ least squares problem, solve $k\times d$ problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \|\boldsymbol{M}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{M}\boldsymbol{b}\|_2^2 \qquad (\boldsymbol{M} \in \mathbb{R}^{k\times n})\,.$$

- Works well when $\boldsymbol{M}$ satisfies subspace J-L property for CS($\boldsymbol{A}$).

Low rank approximation

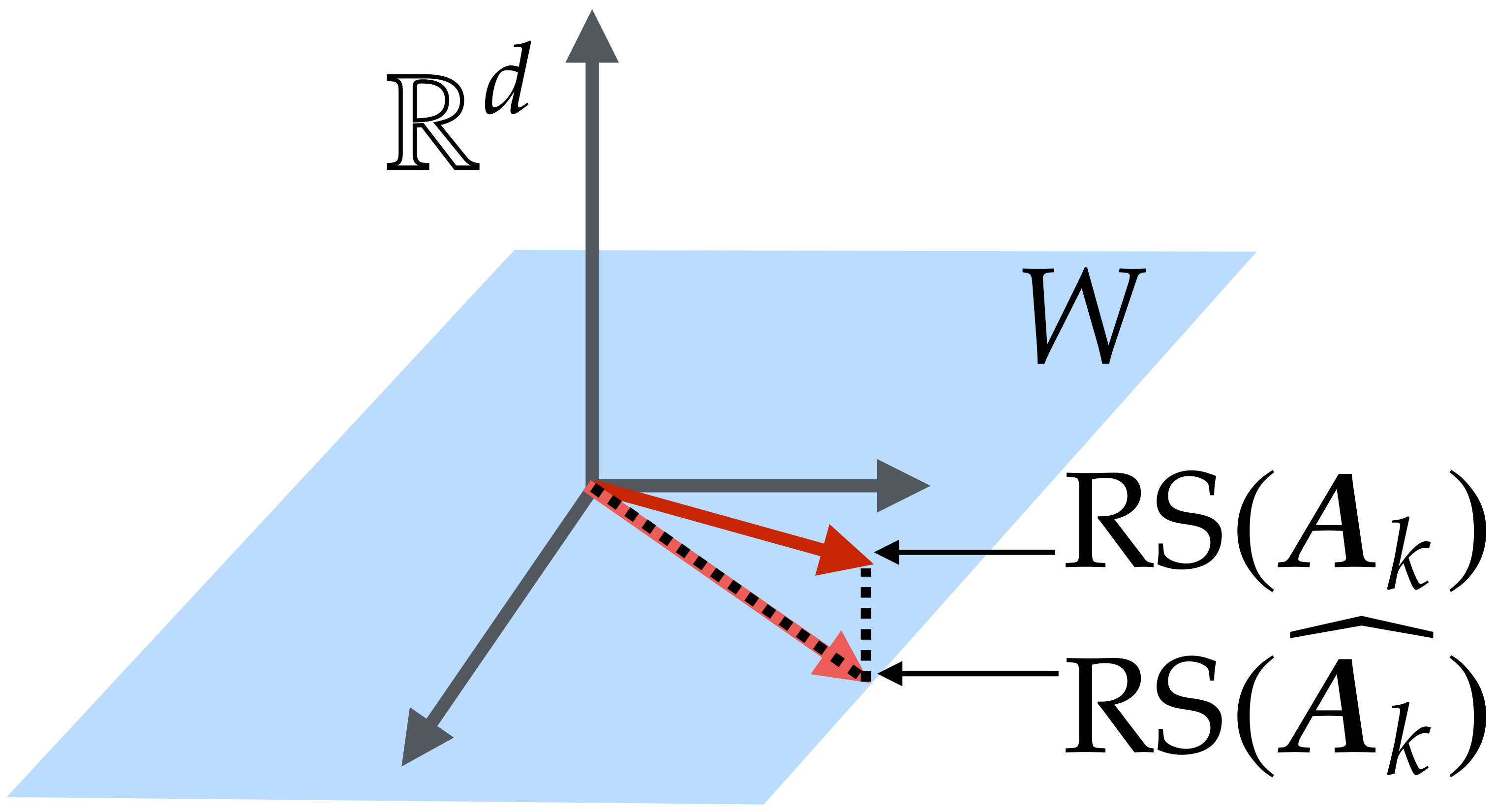
- Optimal rank-k approximation $\boldsymbol{A}_k$ to $\boldsymbol{A}$ can be obtained via SVD.
  - For $n \times d$ matrices, SVD takes time $O(nd^2)$.
- But we only care about rank-k approximation ($k \ll d$), not full SVD.
  - Can we do this faster than the full SVD?

Sketch-and-solve (Sarlós, 2006)

1. Identify a subspace

$$W \subseteq \mathbb{R}^d, \qquad \dim(W) = \ell \geq k$$

containing row space of a near-optimal rank-k approximation to $\boldsymbol{A}$.



2. Extract this near optimal rank-k approximation to $\boldsymbol{A}$ from W.

Capturing a low-rank approximation

- **Task 1**: Find a subspace W containing the row space of a near-optimal rank-k approximation to $\boldsymbol{A}$.

- **Some options**:
  1. Row space of $\boldsymbol{A}$.
     + Contains the row space of $\boldsymbol{A}_k$, the optimal rank-k approximation.
     - But involves working directly with $\boldsymbol{A}$, which is $n \times d$.
  2. Span of $O(k)$ rows of $\boldsymbol{A}$ (perhaps chosen uniformly at random).
     + Small size representation ($O(k) \times d$).
     - But might not contain row space of a good rank-k approximation to $\boldsymbol{A}$.
  3. Span of $O(k)$ random linear combinations of $\boldsymbol{A}$'s rows.

Random linear combinations

Theorem 1. For any $\ell \in \mathbb{N}$ with $\ell \geq \frac{Ck}{\varepsilon}$, there is a linear map $\boldsymbol{M} \in \mathbb{R}^{\ell \times n}$ s.t.

$$\min_{\substack{\boldsymbol{X} \in \mathbb{R}^{n \times \ell}: \\ \operatorname{rank}(\boldsymbol{X}) \leq k}} \|\boldsymbol{A} - \boldsymbol{X}\boldsymbol{M}\boldsymbol{A}\|_{\mathrm{F}}^2 \;\leq\; (1+\varepsilon)\|\boldsymbol{A} - \boldsymbol{A}_k\|_{\mathrm{F}}^2 \,.$$

In fact, the random linear maps used to prove the J-L lemma satisfy this property with high probability.

Finding the low rank approximation

- **Task 2**: Find the rank-k approximation whose row space is in W.

- **Direct approach**:
  1. Get ONB for W.
  2. Project rows of $\boldsymbol{A}$ to W.
  3. Do PCA on projected rows, get ONB for best k-dim. subspace W'.
  4. Project rows of $\boldsymbol{A}$ to W'.

Theorem 2. $$\|\boldsymbol{A} - \boldsymbol{A}\boldsymbol{\Pi}_{W'}\|_{\mathrm{F}}^2 = \min_{\substack{\boldsymbol{X} \in \mathbb{R}^{n \times \ell}: \\ \mathrm{rank}(\boldsymbol{X}) \leq k}} \|\boldsymbol{A} - \boldsymbol{X}\boldsymbol{M}\boldsymbol{A}\|_{\mathrm{F}}^2$$

Actual computational steps

1. Use Gram-Schmidt to get orthonormal basis for row space of $\boldsymbol{MA}$: $d \times \ell$ matrix $\boldsymbol{Q}$.

2. Get coordinates of orthogonal projection of rows of $\boldsymbol{A}$ onto W: $n \times \ell$ matrix $\boldsymbol{AQ}$.

3. Compute SVD of $\boldsymbol{AQ}$ to get matrix of top-k right singular vectors: $\ell \times k$ matrix $\boldsymbol{V}$.

4. Get coordinates of orthogonal projection of rows of $\boldsymbol{A}$ onto $W' = \mathrm{span}(\boldsymbol{QV})$: $n \times k$ matrix $\boldsymbol{AQV}$.

Theorem 2. $$\|\boldsymbol{A} - \boldsymbol{A}(\boldsymbol{QV})(\boldsymbol{QV})^\top\|_\mathrm{F}^2 = \min_{\substack{\boldsymbol{X} \in \mathbb{R}^{n \times \ell}: \\ \mathrm{rank}(\boldsymbol{X}) \leq k}} \|\boldsymbol{A} - \boldsymbol{XMA}\|_\mathrm{F}^2$$

Proof of Theorem 2

- **To show**: if $\boldsymbol{B}$ is best rank-k approximation to $\boldsymbol{AQ}$, then $\boldsymbol{BQ}^\top$ is best rank-k approximation to $\boldsymbol{A}$ among rank-k matrices with row space contained in $\mathrm{RS}(\boldsymbol{Q}^\top)$.

Random linear combinations

Theorem 1. For any $\ell \in \mathbb{N}$ with $\ell \geq \frac{Ck}{\varepsilon}$, there is a linear map $\boldsymbol{M} \in \mathbb{R}^{\ell \times n}$ s.t.

$$\min_{\substack{\boldsymbol{X} \in \mathbb{R}^{n \times \ell}: \\ \operatorname{rank}(\boldsymbol{X}) \leq k}} \|\boldsymbol{A} - \boldsymbol{X}\boldsymbol{M}\boldsymbol{A}\|_{\mathrm{F}}^2 \leq (1+\varepsilon)\|\boldsymbol{A} - \boldsymbol{A}_k\|_{\mathrm{F}}^2 .$$

In fact, the random linear maps used to prove the J-L lemma satisfy this property with high probability.

Proof of Theorem 1

- **To show**: $\mathrm{RS}(\boldsymbol{MA})$ contains a good rank-k approximation to $\boldsymbol{A}$.

- Notation:
  $\boldsymbol{U} = [\ \boldsymbol{u}_1 \mid \boldsymbol{u}_2 \mid \cdots \mid \boldsymbol{u}_k\ ]$ top-k left singular vectors of $\boldsymbol{A}$.
  $\boldsymbol{A} - \boldsymbol{A}_k =: [\ \boldsymbol{r}_1 \mid \boldsymbol{r}_2 \mid \cdots \mid \boldsymbol{r}_d\ ]$ residual vectors, orthogonal to the $\boldsymbol{u}_i$.

- **As a warm-up**: first consider the easy case where $\mathrm{rank}(\boldsymbol{A}) = k$.

  - Need to ensure that $\mathrm{RS}(\boldsymbol{A}_k)$ is a subspace of $\mathrm{RS}(\boldsymbol{MA}) = \mathrm{RS}(\boldsymbol{MA}_k)$.

  - Equivalent to $\mathrm{rank}(\boldsymbol{MU}) = k$ (i.e., invertibility of $(\boldsymbol{MU})^\top \boldsymbol{MU}$).

Sufficient properties of $\boldsymbol{M}$

1. $\boldsymbol{M}$ is a subspace embedding for CS($\boldsymbol{U}$) with constant distortion:

$$\|\boldsymbol{M}\boldsymbol{x}\|_2^2 \;\geq\; \frac{1}{2}\|\boldsymbol{x}\|_2^2 \quad \forall \boldsymbol{x} \in \mathrm{CS}(\boldsymbol{U})\,.$$

2. $\boldsymbol{M}$ approximately preserves orthogonality of $\boldsymbol{u}_i$ and $\boldsymbol{r}_j$:

$$\langle \boldsymbol{M}\boldsymbol{u}_i, \boldsymbol{M}\boldsymbol{r}_j\rangle^2 \;\leq\; \frac{\varepsilon}{4k}\|\boldsymbol{r}_j\|_2^2 \quad \forall i, j\,.$$

Easy to show existence of $\boldsymbol{M} \in \mathbb{R}^{\ell\times n}$ with $\ell \geq \frac{Ck\log d}{\varepsilon}$.

Proof of Theorem 1 (cont.)

- A rank-k approximation whose row space is in RS($\boldsymbol{MA}$):

$$\boldsymbol{U}(\boldsymbol{MU})^{\dagger}\boldsymbol{MA}$$

  - **Note**: when Property 1 holds, and rank($\boldsymbol{A}$) = k, this is exactly $\boldsymbol{A}_k$.

- **To show**: for any matrix $\boldsymbol{M}$ satisfying Property 1 and Property 2,

$$\|\boldsymbol{A} - \boldsymbol{U}(\boldsymbol{MU})^{\dagger}\boldsymbol{MA}\|_{\mathrm{F}}^2 \ \leq \ (1+\varepsilon)\|\boldsymbol{A} - \boldsymbol{A}_k\|_{\mathrm{F}}^2$$

Proof of Theorem 1 (cont.)

- Express

$$\|\boldsymbol{A} - \boldsymbol{U}(\boldsymbol{M}\boldsymbol{U})^{\dagger}\boldsymbol{M}\boldsymbol{A}\|_{\mathrm{F}}^{2}$$

  in terms of

$$\|\boldsymbol{A} - \boldsymbol{A}_k\|_{\mathrm{F}}^{2} \quad \text{and} \quad \sum_{i=1}^{k}\sum_{j=1}^{d}\langle \boldsymbol{M}\boldsymbol{u}_i, \boldsymbol{M}\boldsymbol{r}_j\rangle^2$$