CS 598 (CRN 62819) Topics in Algorithms, Spring 2015

Homework # 2 **Due: February 18, 2015**

Homework is due in class (on paper) on the due date.

1. Make a Piazza comment related to Chapter 2 if you have not done so yet.

2. Make a Piazza comment related to Chapter 3 if you have not done so yet.

3. Suppose matrix A is a database of restaurant ratings: each row is a person, each column is a restaurant, and a_{ij} represents how much person i likes restaurant j. What might $\mathbf{v}_1$ represent? What about $\mathbf{u}_1$? How about the gap $\sigma_1 - \sigma_2$? (There are multiple reasonable answers here).

   For this specific question, feel free to combine with question 2 and either post your thoughts or comment on someone else's thoughts, or post thoughts about what $\mathbf{v}_2, \mathbf{u}_2$ might represent, etc. You could also read up and make a post about how one might do SVD when A has missing entries (not all people have ranked all the restaurants).

4. Consider the matrix

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 2 \\ 1 & -2 \\ -1 & -2 \end{bmatrix}$$

   (a) Run the power method starting from $x = \binom{1}{1}$ for $k = 3$ steps. What does this give as an estimate of $\mathbf{v}_1$?

   (b) What actually are the $\mathbf{v}_i$'s, σ_i's, and $\mathbf{u}_i$'s? It may be easiest to do this by computing the eigenvectors of $B = A^T A$.

5. Prove that $\mathbf{u}_1$ is the first singular vector for A^T (let's assume $\sigma_1 > \sigma_2$ so we can say "the" instead of "a"). Hint: we know the power method applied to $C = AA^T$ will approach the first singular vector of A^T if it approaches any fixed vector at all.

6. Consider a set S of n points in R^d whose center of mass is the origin. As we have been discussing in class, the first singular vector (for a matrix in which each of these points is a row) gives the line through the origin that minimizes the sum of squared distances of all the points in S to that line. What if we allowed lines not through the origin? Here we will prove that such lines don't help. Specifically, for any line ℓ, the line ℓ' parallel to ℓ that goes through the origin is at least as good.

   Formally, fix some line ℓ. Let $\mathbf{b}$ be the point on ℓ closest to the origin and let $\mathbf{v}$ be a unit length vector parallel to ℓ. Mathematically, we have $\ell = \{\mathbf{b} + \lambda\mathbf{v} : \lambda \in R\}$ (convince yourself of this fact before going on). Let $\ell' = \{\lambda\mathbf{v} : \lambda \in R\}$ be the line parallel to ℓ that goes through the origin.

(a) Explain why for any point $\mathbf{x}$, its squared distance to ℓ' is $||\mathbf{x}||^2 - (\mathbf{v} \cdot \mathbf{x})^2$.

(b) What is the analogous formula for distance squared of a point $\mathbf{x}$ to ℓ?

(c) Now prove the theorem: prove that the sum of squared distances of points in S to ℓ' is less than or equal to the sum of squared distances of points in S to ℓ. Be clear where you use the fact that the center of mass of S is the origin.