Colorado CSCI 5454: Algorithms
Homework 9

Instructor: Bo Waggoner
Due: Friday, November 22, 2019 at 11:59pm
Turn in electronically via Gradescope.

Remember to list the people you worked with and any outside sources used (if none, write "none").

Problem 1 (12 points)

Suppose you have a database of movie recommendations stored as a matrix $A \in \mathbb{R}^{n \times d}$, where $A(i,j)$ is person i's rating of movie j, a real number between zero and one.

There are n people and d movies, so A has n rows and d columns.

Now you take the singular value decomposition,

$$A = U \; D \; V^{\intercal}$$

where:

- $U \in \mathbb{R}^{n \times n}$ is an orthogonal matrix[1].
- $D \in \mathbb{R}^{n \times d}$ is a diagonal matrix[2], and the entries are sorted from largest to smallest, and only r entries are nonzero where r is the rank of A.
- $V \in \mathbb{R}^{d \times d}$ is an orthogonal matrix, and $V^{\intercal}$ is its transpose.

Recall that the columns $u_1, \dots, u_n$ of U are the *left singular vectors*, the diagonal entries $\sigma_1, \dots, \sigma_r$ of D are the *singular values*, and the columns $v_1, \dots, v_d$ of V are the *right* singular vectors.

Recall that we obtain a rank-k approximation by taking the first k columns of U and V, along with the first k rows and columns of D. This gives us $U_k \in \mathbb{R}^{n \times k}$, $D_k \in \mathbb{R}^{k \times k}$, and $V_k \in \mathbb{R}^{d \times k}$, with

$$A_k := U_k \; D_k \; V_k^{\intercal}.$$

In this problem, we'll show that A_k is a sum of k rank-one matrices, from "most important" to "least". We will also bound how much accuracy is lost by dropping the "unimportant" matrices from the sum.

Part a (4 points) Show that $A_k(j, \ell) = \sum_{i=1}^{k} \sigma_i u_i(j) v_i(\ell)$.

[1] Every row is a unit vector, and the rows are all pairwise orthogonal; and this is also true of the columns.
[2] All entries are zero except the (i,i) entries.

Part b (2 points) Recall that if u and v are vectors, then their outer product $uv^\intercal$ is a matrix whose (j, ℓ) entry is $u(j)v(\ell)$. Using this and the previous part, show the following identity:

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\intercal$$

That is, A_k is the sum of these k matrices, one for each set of singular value+vectors.

Part c (2 points) Suppose you have computed the SVD and listed all the singular values and vectors of A. You used these to compute the approximation A_k. Using the previous part, what is a quick way to now compute the better approximation A_{k+1}?

Part d (2 points) When we approximate A by A_k, let the *remainder* be $R_k := A - A_k$. Show that $R_k(j, \ell) = \sum_{i=k+1}^{r} \sigma_i u_i(j) v_i(\ell)$.

Part e (2 points) Use the previous part to argue that $\|A - A_k\|_F^2 \leq \sum_{i=k+1}^{r} \sigma_i^2$. In other words, the total error in the approximation of A_k is bounded by the small singular values that are dropped.

(Recall that for a matrix R, $\|R\|_F^2 := \sum_{j,\ell} R(j, \ell)^2$. Also recall that each u_i and v_i are unit vectors.)

Problem 2 (4 points)

Your friend is boasting about the following construction. "In m-dimensional space," she says, "I put a point at $(\frac{1}{\sqrt{2}}, 0, \ldots, 0)$. Then I put one at $(0, \frac{1}{\sqrt{2}}, 0, \ldots, 0)$. And so on. Eventually, I have placed m points in just m dimensions, such that the distance between any pair of points is exactly one!"

"That's nothing," you say. "I can place m points with all pairwise distances at between 0.9 and 1.1, and I only need O (______________) dimensions!"

(Fill in the blank and carefully justify your answer.)