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 *(updated)*, when r is the rank of A:

- $U \in \mathbb{R}^{n \times r}$ is an orthogonal matrix[1].
- $D \in \mathbb{R}^{r \times r}$ is a diagonal matrix[2], and the entries are sorted from largest to smallest.
- $V \in \mathbb{R}^{d \times r}$ 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 of D. This gives us $U_k \in \mathbb{R}^{n \times k}$, $D_k \in \mathbb{R}^{k \times d}$, 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.

Solution. Let $B = U_k D_k$. Then by the rules of matrix multiplication, $B(j,i) = \sum_{o=1}^{k} U_k(j,o) D_k(o,i)$. Note $U_k(j,o) = u_o(j)$, where u_o is the oth column. Since $D_k(o,i) = 0$ unless $o = i$, in which case it is σ_i, we have $B(j,i) = u_i(j)\sigma_i$.

Then $A_k = BV_k^{\intercal}$. Note $V_k^{\intercal}(i,\ell) = V_k(\ell,i) = v_i(\ell)$ where v_i is the ith column of V_k. So

$$\begin{aligned} A_k(j,\ell) &= \sum_{i=1}^{k} B(j,i) V_k^{\intercal}(i,\ell) \\ &= \sum_{i=1}^{k} B(j,i) v_i(\ell) \\ &= \sum_{i=1}^{k} u_i(j)\sigma_i v_i(\ell). \end{aligned}$$

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.

Solution. Let us define $C = \sum_{i=1}^{k} \sigma_I u_i v_i^{\intercal}$. We just need to show $C(j,\ell) = A_k(j,\ell)$ for all entries j,ℓ. If we look at the matrix $u_i v_i^{\intercal}$, then its j,ℓ entry is $u_i(j)v_i(\ell)$. So $C(j,\ell) = \sum_{i=1}^{k} \sigma_i u_i(j) v_i(\ell)$, which equals $A_k(j,\ell)$ by the previous part.

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}?

Solution. By the previous part, we can write $A_{k+1} = A_k + \sigma_{k+1} u_{k+1} v_{k+1}^{\intercal}$. So we just need to add this rank-one matrix to A_k.

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

Solution. If we set $k = r$ in the previous parts, we get all of A, i.e. $A = \sum_{i=1}^{r} \sigma_i u_i v_i^{\intercal}$. So $A - A_k = \sum_{i=k+1}^{r} \sigma_i u_i v_i^{\intercal}$, which implies the claim.

Part e (2 points) Use the previous part to argue that $\|A - A_k\|_F^2 \le \sum_{i=k+1}^r \sigma_i^2$. 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.)

Solution. We also need to recall that u_i is orthogonal to all $u_{i'}$ for $i \neq i'$, and similarly the columns of V are orthogonal. So we can take a "brute-force" approach of expanding it out:

$$\begin{aligned}
&\sum_{j,\ell} R_k(j,\ell)^2 \\
&= \sum_{j,\ell} \sum_{i=k+1}^{r} \sum_{i'=k+1}^{r} \sigma_i \sigma_{i'} u_i(j) u_{i'}(j) v_i(\ell) v_{i'}(\ell) \\
&= \sum_{i=k+1}^{r} \sum_{i'=k+1}^{r} \sigma_i \sigma_{i'} \sum_j u_i(j) u_{i'}(j) \sum_\ell v_i(\ell) v_{i'}(\ell) \\
&= \sum_{i=k+1}^{r} \sigma_i^2 .
\end{aligned}$$

The key is that $\sum_\ell v_i(\ell) v_{i'}(\ell) = v_i \cdot v_{i'}$, so it is zero if $i \neq i'$, otherwise it is one. The same goes for u_i and $u_{i'}$.

Note we actually get equality.

Problem 2 (4 points)

Your friend is boasting about the following construction. "In m-dimensional space," she says, "I put a point at $(\frac{1}{\sqrt{2}}, 0, \dots, 0)$. Then I put one at $(0, \frac{1}{\sqrt{2}}, 0, \dots, 0)$. 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.)

Solution. We can put $\ln(m)$ in the blank.

There are m points in our friend's construction. The Johnson-Lindenstrauss lemma says that, if we start from our friend's construction, we can take $\epsilon = 0.1$ and any $\delta < 1$ and get, with probability at least $1 - \delta$, all the projected points will have pairwise distance in $[0.9, 1.1]$, for any choice of dimension $d \ge \frac{8}{\epsilon^2} \ln\left(\frac{m}{\delta}\right) = 800 \ln(m/\delta)$.

Now, this holds for any δ. So if we pick any $\delta < 1$, there is positive probability of being able to place the points in this manner. So there will be some way to place the points in any dimension $d > 800 \ln(m)$.