A1 (Extra Credit)

Graded

Student

Sugi Venugeethan

Total Points

5 / 5 pts

Question 1

4a solution

1 / 1 pt

✓ + 1 pt $(\sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T)_{k\ell} = \sum_{i=1}^r \sigma_i (\mathbf{u}_i \mathbf{v}_i^T)_{k\ell}$
$= \sum_{i=1}^r \sigma_i (\mathbf{u}_i)_k (\mathbf{v}_i)_\ell$
$= \sum_{i=1}^r \sigma_i \mathbf{V}_{\ell i}^T \mathbf{U}_{ik}$
$= \sum_{i=1}^r \mathbf{V}_{\ell i}^T (\mathbf{D}\mathbf{U})_{ik}$
$= (\mathbf{V}^T \mathbf{D}\mathbf{U})_{\ell k}$
$= (\mathbf{U}\mathbf{D}\mathbf{V}^T)_{k\ell}$
$= \mathbf{A}_{k\ell}$

Question 2

4b solution

1 / 1 pt

✓ + 1 pt Recalling that the $\mathbf{v}_i$ are orthonormal, we can use the result from (a) to see that

$$\mathbf{A}\mathbf{v}_j = \left(\sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T\right)\mathbf{v}_j$$
$$= \left(\sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \mathbf{v}_j\right)$$
$$= \sigma_j \mathbf{u}_j$$

so $\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$ as desired.

Question 3

4c solution

1 / 1 pt

✓ + 1 pt For an arbitrary vector $\mathbf{a}$, using that the $\mathbf{v}_i$ are orthonormal (see the hint for why this is important), we have that the projection of $\mathbf{a}$ onto $V_k = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is $\sum_{i=1}^k (\mathbf{a}^T \mathbf{v}_i)\mathbf{v}_i$ from the fact that

$$\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) = \left(\frac{\mathbf{v}^T\mathbf{w}}{||\mathbf{w}||_2^2}\right)\mathbf{w} = (\mathbf{v}^T\hat{\mathbf{w}})\hat{\mathbf{w}}$$

Thus the matrix whose rows are the projections of each row of $\mathbf{A}$ onto V_k is given by

$$\sum_{i=1}^k \mathbf{A}\mathbf{v}_i\mathbf{v}_i^T$$

Using the result from (b) we have that $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ so

$$\sum_{i=1}^k \mathbf{A}\mathbf{v}_i\mathbf{v}_i^T = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \mathbf{A}_k$$

as desired.

Question 4

4d solution

2 / 2 pts

+ 2 pts For a matrix $\mathbf{M}$ and linear subspace V, let $\mathrm{proj}_V(\mathbf{M})$ represent the matrix with rows $\mathrm{proj}_V(m_i)$, where m_i is the ith row of $\mathbf{M}$.

Consider an arbitrary matrix $\mathbf{B} \in \mathbb{R}^{n \times d}$ of rank k. Then let the k-dimensional space W be the span of the rows of $\mathbf{B}$, so W has dimension k. We see that:

$$||\mathbf{A} - \mathbf{B}||_F^2 = \sum_{i=1}^n ||a_i - b_i||_2^2$$

where a_i, b_i are the rows of $\mathbf{A}$ and $\mathbf{B}$. From the fact that

$$\mathrm{proj}_W(\mathbf{v}) := \arg\min_{w \in W} ||w - \mathbf{v}||_2^2,$$

we know that

$$||\mathbf{A} - \mathbf{B}||_F^2 = \sum_{i=1}^n ||a_i - b_i||_2^2 \geq \sum_{i=1}^n ||a_i - \mathrm{proj}_W(a_i)||_2^2 = ||\mathbf{A} - \mathrm{proj}_W(\mathbf{A})||_F^2$$

But since V_k is the best-fit k-dimensional subspace for the rows of $\mathbf{A}$, we know that

$$||\mathbf{A} - \mathrm{proj}_{V_k}(\mathbf{A})||_F^2 = \sum_{i=1}^n ||a_i - \mathrm{proj}_{V_k}(a_i)||_2^2 \leq \sum_{i=1}^n ||a_i - \mathrm{proj}_W(a_i)||_2^2 = ||\mathbf{A} - \mathrm{proj}_W(\mathbf{A})||_F^2$$

Part (c) tells us that $\mathrm{proj}_{V_k}(\mathbf{A}) = \mathbf{A}_k$. Putting everything together gives

$$||\mathbf{A} - \mathbf{A}_k||_F^2 \leq ||\mathbf{A} - \mathrm{proj}_W(\mathbf{A})||_F^2 \leq ||\mathbf{A} - \mathbf{B}||_F^2$$

Since the above inequality is true for any rank k matrix $\mathbf{B}$, it follows that

$$\mathbf{A}_k = \arg\min_{\mathrm{rank}(\mathbf{B})=k} ||\mathbf{A} - \mathbf{B}||_F^2 = \arg\min_{\mathrm{rank}(\mathbf{B})=k} ||\mathbf{A} - \mathbf{B}||_F$$

Question assigned to the following page: 1

This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the README.md for this assignment includes instructions to regenerate this handout with your typeset LaTeX solutions.

4.a Suppose we have a matrix $A \in R^{n \times d}$ with SVD $A = UDV^T$, where $U \in R^{n \times r}$, $D \in R^{r \times r}$, $V \in R^{d \times r}$, show that $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

PROOF:

Matrix A can transform vectors v_i. Now vector v can be written as a linear combination of $v_1, v_2, \ldots v_r$ and a vector perpendicular to v_i. A_v is linear combination of $Av_1, Av_2 \ldots Av_r$. $Av_1, Av_2, Av_3 \ldots Av_r$ are set of vectors associated with A. If we normalize to length one, we get

$$\boxed{u_i = \frac{1}{\sigma_i(A)} Av_i} \quad — (1)$$

Vectors u_1, u_2, u_r are called left singular vectors. $v_1, v_2 \ldots v_r$ are called right singular vectors.

A is $n \times d$ matrix with singular vectors $v_1, v_2 \ldots v_r$ and corresponding singular values $\sigma_1, \sigma_2 \ldots \sigma_r$. From left singular vectors u_i as per (1) above, A can be decomposed into sum of rank one matrices

$$\boxed{A = \sum_{i=1}^{r} \sigma_i u_i v_i^T}$$

Note: Matrices A and B are identical if and only if for all vectors v, $Av = Bv$

For each singular vector v_j, $Av_j = \sum_{i=1}^{r} \sigma_i u_i v_i^T v_j$. Since any vector v can be expressed as linear combination of singular vectors plus a vector perpendicular to v_i $Av = \sum_{i=1}^{r} \sigma_i u_i v_i^T v$

$\therefore A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

Question assigned to the following page: 2

4.b Show that $u_i = \frac{1}{\sigma_i} A v_i$. In particular, the components of u_i represent the size of projection of the rows of A onto v_i (scaled by σ_i)

PROOF:

Consider rows of A as n points in a d-dimensional space. Let us now consider the best fit line through origin. Let v be a unit vector along this line. Now, the length of projection of a_i, the ith row of A, onto v is $|a_i \cdot v|$.

Sum of length squared of projections is $|Av|^2$. The best fit line is one maximizing $|Av|^2$ and hence minimizing sum of squared distance of points to the line.

$\sigma_1(A) = |Av_1|$ is first singular value of A

$\sigma_2(A) = |Av_2|$ is second singular value of A

A (matrix) can transform vector v_i. Every vector v can be written as linear combination of $v_1, v_2 \dots v_r$ and a vector perpendicular to all v_i.

Av_1, Av_2 from above form a fundamental set of vectors associated with A.

we get left singular vectors of A, $u_1, u_2 \dots u_r$ by normalizing Av_i vector to length one

$\therefore \boxed{u_i = \frac{1}{\sigma_i} A v_i}$

In other words, if vs are orthogonal vectors, us are transformed orthogonal vectors.

Question assigned to the following page: 3

4.c

Truncated SVD

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

To show:
Rows of A_k are the projections of rows of A onto the subspace of V_k spanned by first k right singular vector

Let us say a be an arbitrary row vector. v_i are orthonormal, so the projection of the vector a onto V_k is given by

$$\sum_{i=1}^{k} (a \cdot v_i) v_i^T$$

∴ The matrix whose rows are the projections of the rows of A onto V_k is given by

$$\sum_{i=1}^{k} A v_i v_i^T$$

Substituting for Av_i, we get

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

∴ Matrix A_k is best rank k approximation to A.

Question assigned to the following page: 4

4.d

The frobenius norm of matrix $M \in R^{n \times m}$

$$\|M\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} M_{ij}^2}$$

To show $A_k = \underset{\text{rank}(B)=k}{\text{argmin}} \|A - B\|_F$

V_k is best fit k-dimensional subspace for rows of A.

Each row of B is the projection of corresponding row of A, it follows that $\|A - B\|_F^2$ is sum of squared distances of rows of A to V. V be space spanned by rows of B.

Since A_k minimizes the sum of squared distance of rows of A to k-dimensional subspace it can be said that

$$\boxed{A_k = \underset{\text{rank}(B)=k}{\text{argmin}} \|A - B\|_F}$$