Matrix Theory (AI1000): Practice Problems

Singular Value Decomposition and Principal Component Analysis

---

Part I: Computing the SVD

P1. Let $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$.

(a) Compute A^TA and find its eigenvalues and orthonormal eigenvectors $\mathbf{v}_1, \mathbf{v}_2$. Hence determine the singular values $\sigma_1 \geq \sigma_2$ of A.

(b) Compute the left singular vectors via $\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$. Verify that $U = [\mathbf{u}_1 \ \mathbf{u}_2]$ is orthogonal.

(c) Write the full SVD $A = U\Sigma V^T$ and verify it by multiplying out.

(d) Give a geometric interpretation: what does A do to the unit circle in $\mathbb{R}^2$? Identify the semi-axes of the resulting ellipse.

P2. Let $B = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$.

(a) Compute $BB^T \in \mathbb{R}^{2\times 2}$ and find its eigenvalues and orthonormal eigenvectors. Hence determine the singular values of B.

(b) Compute $B^TB \in \mathbb{R}^{3\times 3}$, find its eigenvalues, and obtain the right singular vectors $\mathbf{v}_1, \mathbf{v}_2$ from the nonzero eigenvalues. Extend to a full orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for $\mathbb{R}^3$ by finding $\mathbf{v}_3 \in \text{Null}(B)$.

(c) Write the full SVD $B = U\Sigma V^T$ with $U \in \mathbb{R}^{2\times 2}$, $\Sigma \in \mathbb{R}^{2\times 3}$, and $V \in \mathbb{R}^{3\times 3}$.

(d) Verify that the outer-product expansion $B = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T$ recovers B.

P3. Let $C = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

(a) Find the eigenvalues $\lambda_1 \geq \lambda_2$ and corresponding orthonormal eigenvectors of C.

(b) C is symmetric positive definite. Explain, in general, how the SVD of a symmetric positive definite matrix S relates to its eigendecomposition $S = Q\Lambda Q^T$. What are the singular values and what are U, Σ, V in terms of Q and Λ?

(c) Now suppose $D = \begin{pmatrix} 2 & 3 \\ 3 & -2 \end{pmatrix}$ (indefinite symmetric). Find the SVD of D. Explain why the SVD and eigendecomposition of D differ, and how the signs of the negative eigenvalues are handled.

P4. Let $\mathbf{u} = \frac{1}{\sqrt{5}}\begin{pmatrix}1\\2\end{pmatrix}$, $\mathbf{v} = \frac{1}{\sqrt{10}}\begin{pmatrix}1\\3\end{pmatrix}$, and define $M = 6\,\mathbf{u}\mathbf{v}^T \in \mathbb{R}^{2\times 2}$.

(a) Write down the SVD of M directly (without computing M^TM), explaining your reasoning. What are the singular values, and what are U and V?

(b) Compute M explicitly and verify your SVD.

(c) Let $E = M + \epsilon I$ for small $\epsilon > 0$. Without computing the full SVD of E, explain qualitatively how the singular values and singular vectors of E differ from those of M as $\epsilon \to 0$. (*Hint:* Consider the rank of E versus M, and the effect on M^TM.)

Part II: Properties and Theory of SVD

P5. Let $A \in \mathbb{R}^{m\times n}$ have SVD $A = U\Sigma V^T$ with $r = \operatorname{rank}(A)$. Partition $U = [U_1\ U_2]$ where U_1 contains the first r columns, and similarly $V = [V_1\ V_2]$.

(a) Prove that $\operatorname{Col}(A) = \operatorname{Col}(U_1)$ and $\operatorname{Null}(A^T) = \operatorname{Col}(U_2)$.

(b) Prove that $\operatorname{Row}(A) = \operatorname{Col}(V_1)$ and $\operatorname{Null}(A) = \operatorname{Col}(V_2)$.

(c) Conclude that the four fundamental subspaces of A are orthogonal complements in pairs: $\operatorname{Col}(A) \perp \operatorname{Null}(A^T)$ and $\operatorname{Row}(A) \perp \operatorname{Null}(A)$.

(d) Let $A = \begin{pmatrix}1 & 2 & 0\\ 2 & 4 & 1\\ 0 & 0 & 1\end{pmatrix}$. Without computing the full SVD, find $\operatorname{rank}(A)$ and use the result of parts (a)–(c) to state the dimensions of all four fundamental subspaces.

P6. Let $A \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

(a) Show that the *spectral norm* (operator 2-norm) of A satisfies $\|A\|_2 = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \sigma_1$. (*Hint:* Write $\mathbf{x}$ in the basis of right singular vectors.)

(b) Show that the *Frobenius norm* satisfies $\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2$. (*Hint:* Use $\|A\|_F^2 = \operatorname{tr}(A^TA)$.)

(c) For $A = \begin{pmatrix}1 & 2\\ 2 & 1\\ 1 & 1\end{pmatrix}$, compute $\|A\|_2$ and $\|A\|_F$ using the SVD. (You will need to find the singular values of A.)

(d) Define the *condition number* of A as $\kappa(A) = \sigma_1/\sigma_r$ (for square invertible A). Explain what a large condition number means for solving $A\mathbf{x} = \mathbf{b}$ and relate this to the near-singularity of A.

P7. The *Moore–Penrose pseudoinverse* of $A \in \mathbb{R}^{m\times n}$ with SVD $A = U\Sigma V^T$ is defined as $A^+ = V\Sigma^+ U^T$, where Σ^+ is obtained from Σ by replacing each nonzero singular value σ_i by $1/\sigma_i$ and transposing.

(a) Verify that A^+ satisfies the four Moore–Penrose conditions: (i) $AA^+A = A$; (ii) $A^+AA^+ = A^+$; (iii) $(AA^+)^T = AA^+$; (iv) $(A^+A)^T = A^+A$.

(b) Show that when A has full column rank, $A^+ = (A^TA)^{-1}A^T$ (i.e. the pseudoinverse reduces to the left inverse).

(c) Compute A^+ for $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}$.

(d) Show that $\hat{\mathbf{x}} = A^+\mathbf{b}$ gives the minimum-norm least-squares solution to $A\mathbf{x} = \mathbf{b}$, i.e. it minimises $\|A\mathbf{x} - \mathbf{b}\|$ and among all minimisers has the smallest $\|\mathbf{x}\|$.

Part III: Low-Rank Approximation and the Eckart–Young Theorem

P8. Let $A \in \mathbb{R}^{m\times n}$ have SVD $A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Define the rank-$k$ truncation $A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$.

(a) State the Eckart–Young–Mirsky theorem precisely for both the spectral norm and the Frobenius norm.

(b) Prove the Frobenius-norm version: for any matrix B with $\operatorname{rank}(B) \leq k$, $\|A - A_k\|_F \leq \|A - B\|_F$. (*Hint:* Use the unitary invariance of the Frobenius norm and the fact that the best rank-k approximation to a diagonal matrix is obtained by zeroing out all but the k largest diagonal entries.)

(c) Let $A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$. Compute the full SVD of A and the best rank-1 approximation A_1. Compute $\|A - A_1\|_F$ and $\|A - A_1\|_2$.

(d) For the matrix A in (c), compute the percentage of the total "energy" (sum of squared singular values) captured by A_1. Is a rank-1 approximation reasonable in this case?

P9. An $m \times n$ grayscale image is represented as a matrix P of pixel intensities. Suppose P has singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$.

(a) The rank-k approximation P_k requires storing $k(m + n + 1)$ numbers (the k singular values, k left singular vectors of length m, and k right singular vectors of length n). The original image requires storing mn numbers. Derive the compression ratio $\rho(k) = mn \,/\, [k(m + n + 1)]$ and find the maximum k for which P_k uses fewer numbers than the original.

(b) For a 512×512 image, what is the maximum k giving compression? What fraction of the singular values does this represent?

(c) Suppose P has singular values $\sigma_i = 100e^{-0.1(i-1)}$ for $i = 1, \ldots, 100$ (and $\sigma_i = 0$ for $i > 100$). Find the smallest k such that the rank-k approximation captures at least 99% of the total energy $\|P\|_F^2 = \sum_i \sigma_i^2$.

(d) Explain why images tend to have rapidly decaying singular values, making low-rank approximation effective in practice. What property of natural images underlies this?

Part IV: Principal Component Analysis

P10. The following data matrix has columns representing 5 observations of a 3-dimensional random variable (already mean-centred):

$$X = \begin{pmatrix} 1 & -1 & 2 & -1 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 5}.$$

(a) Compute the sample covariance matrix $S = \frac{1}{n-1} XX^T \in \mathbb{R}^{3\times 3}$ (where $n = 5$). Verify it is symmetric positive semi-definite.

(b) Find the eigenvalues and eigenvectors of S. These are the *principal values* and *principal directions* of the data.

(c) Compute the fraction of total variance explained by each principal component. How many components are needed to explain at least 90% of the variance?

(d) Project the data onto the first two principal components to obtain the 2×5 score matrix $Z = W^T X$, where W contains the top-2 eigenvectors of S as columns. Interpret the scores geometrically.

P11. This problem establishes the equivalence between PCA and SVD. Let $X \in \mathbb{R}^{p\times n}$ be a mean-centred data matrix (rows are variables, columns are observations).

(a) Show that the sample covariance matrix $S = \frac{1}{n-1} XX^T$ has eigendecomposition $S = Q\Lambda Q^T$, where the columns of Q are the right singular vectors of X (i.e. the left singular vectors of X when X is viewed as a map $\mathbb{R}^n \to \mathbb{R}^p$) and $\Lambda = \frac{1}{n-1}\operatorname{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. Relate the singular values of X to the eigenvalues of S.

(b) Explain why computing PCA via the SVD of X is preferred numerically over forming XX^T explicitly and finding its eigendecomposition. (*Hint:* Consider the condition number of S versus that of X.)

(c) Suppose $X \in \mathbb{R}^{p\times n}$ with $p \gg n$ (many variables, few observations). Which matrix is cheaper to eigendecompose: $XX^T \in \mathbb{R}^{p\times p}$ or $X^T X \in \mathbb{R}^{n\times n}$? Show

that the nonzero eigenvalues of XX^T and X^TX coincide, so the PCA can be computed from the smaller matrix.

(d) In the dual PCA approach using X^TX, suppose $\mathbf{w}$ is an eigenvector of X^TX with eigenvalue $\lambda > 0$. Show that $X\mathbf{w}$ is an eigenvector of XX^T with the same eigenvalue λ. Hence derive the principal directions from $\mathbf{w}$.

P12. A dataset of face images is stored as a matrix $F \in \mathbb{R}^{d\times n}$, where each column is a vectorised (mean-subtracted) image of dimension $d = 1024$ pixels and $n = 200$ images.

(a) Suppose the top k principal components explain 95% of the total variance. Write the formula for the reconstructed image $\hat{\mathbf{f}}_j$ of the j-th face $\mathbf{f}_j$, using the top-k principal directions $\mathbf{w}_1, \dots, \mathbf{w}_k$. What is the reconstruction error $\|\mathbf{f}_j - \hat{\mathbf{f}}_j\|^2$ in terms of the singular values of F?

(b) Explain the concept of "eigenfaces": what do the principal components $\mathbf{w}_1, \dots, \mathbf{w}_k$ represent, and why do they resemble face-like structures?

(c) A new image $\mathbf{g} \in \mathbb{R}^{1024}$ (mean-subtracted) is presented. Describe how to decide whether $\mathbf{g}$ is "face-like" using the PCA basis. What quantity would you threshold, and why?

(d) Suppose two images $\mathbf{f}_i$ and $\mathbf{f}_j$ have score vectors $\mathbf{z}_i, \mathbf{z}_j \in \mathbb{R}^k$. Argue that the Euclidean distance $\|\mathbf{z}_i - \mathbf{z}_j\|$ in the PCA subspace approximates the original distance $\|\mathbf{f}_i - \mathbf{f}_j\|$, and state when the approximation is exact.

P13. (Synthesis) Let $X \in \mathbb{R}^{p\times n}$ be a mean-centred data matrix with SVD $X = U\Sigma V^T$, and let $k < \operatorname{rank}(X)$.

(a) Show that the k-dimensional PCA subspace (spanned by the top k left singular vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$) minimises the total *reconstruction error*

$$\sum_{j=1}^{n} \|\mathbf{x}_j - \hat{\mathbf{x}}_j\|^2,$$

where $\hat{\mathbf{x}}_j$ is the projection of observation $\mathbf{x}_j$ onto the subspace. Relate this to the Eckart–Young theorem.

(b) Dually, show that the k-dimensional PCA subspace maximises the total *projected variance*

$$\sum_{j=1}^{n} \|\hat{\mathbf{x}}_j\|^2 = \|\text{projection of } X \text{ onto subspace}\|_F^2.$$

Why are (a) and (b) equivalent?

(c) Explain the connection between PCA and the least-squares problem. Specifically, if we model the data as $\mathbf{x}_j \approx W\mathbf{z}_j$ where $W \in \mathbb{R}^{p\times k}$ has orthonormal

columns and $\mathbf{z}_j \in \mathbb{R}^k$, and minimise $\sum_j \|\mathbf{x}_j - W\mathbf{z}_j\|^2$ over both W and $\mathbf{z}_j$, show that the optimal W has columns equal to the top k left singular vectors of X.

(d) What goes wrong if we do not mean-centre the data before applying PCA? Give a concrete 2×3 example where failing to mean-centre produces a misleading first principal component.
