AI 391L - FALL 2025

Homework 3: Theory

Question 1

Let A be an $m \times d$ matrix, and let $X = AA^\top$. Assume that X has d distinct, nonzero eigenvalues and that $m \gg d$. Computing the eigendecomposition of X directly is slow since X is $m \times m$. Give an algorithm to find the eigenvectors and eigenvalues of X that only requires computing the eigendecomposition of a $d \times d$ matrix. You may use simple matrix operations and an eigendecomposition "black box," but do not call SVD as a black box.

Solution

Step 1. Form the $d \times d$ matrix $A^\top A$ and compute its eigendecomposition. Let v be an eigenvector of $A^\top A$ with eigenvalue $\lambda > 0$, i.e.,

$$A^\top A\, v = \lambda v.$$

Step 2. Show that Av is an eigenvector of $AA^\top$ with the same eigenvalue:

$$(AA^\top)(Av) = A(A^\top Av) = A(\lambda v) = \lambda(Av).$$

Hence, for every eigenpair (v, λ) of $A^\top A$ with $\lambda > 0$, the vector $u := Av$ is an eigenvector of $AA^\top$ with eigenvalue λ.

Step 3. (Zero eigenvalues.) Any vector in $\mathrm{null}(AA^\top)$ is an eigenvector of $AA^\top$ with eigenvalue 0.

Output. Collect $u_i = Av_i$ for all eigenvectors v_i of $A^\top A$ with eigenvalues $\lambda_i > 0$. Then

$$AA^\top u_i = \lambda_i u_i,$$

so $\{\lambda_i\}$ are the nonzero eigenvalues of X and $\{u_i\}$ are the corresponding eigenvectors. (If desired, normalize u_i to unit length.)

Question 2

In this problem we explore some relationships between SVD, PCA and linear regression.

(a) **[2 points]** True or false: linear regression is primarily a technique of *supervised* learning, i.e. where we are trying to fit a function to labeled data.

(b) **[2 points]** True or false: PCA is primarily a technique of *unsupervised* learning, i.e. where we are trying to find structure in unlabeled data.

(c) **[2 points]** True or false: SVD is primarily an operation on a *dataset* whereas PCA is primarily an operation on a *matrix*.

(d) **[2 points]** A common problem in linear regression is *multicollinearity*, where the input variables are themselves linearly dependent. For example, imagine a healthcare data set where height is measured both in inches and in centimeters. This is a problem because there may now be multiple w satisfying $y = w \cdot x$. Explain how you could use a preprocessing step to solve this problem.

Solution

(a) **True.** Linear regression learns a mapping from inputs to outputs using labeled examples, so it is supervised.

(b) **True.** PCA does not use labels. It just looks for directions in the data that explain the most variance, which is unsupervised.

(c) **False.** SVD is purely a matrix factorization. PCA is a data analysis method that we often carry out by running SVD on a centered data matrix. So PCA is built on top of SVD, not the other way around.

(d) A simple fix is to do PCA first and then run regression on the principal components. This step removes duplicate or dependent features (like inches vs. centimeters) because the components are orthogonal. This way regression has a unique solution and is less sensitive to multicollinearity.