Linear Algebra Working Group :: Day 3

Note: All vector spaces will be finite-dimensional vector spaces over the field $\mathbb{R}$.

1 Principal component analysis and dimensional reduction

Definition 1.1. Given an $m \times N$ matrix X of observations, with columns X_j thought of as m-dimensional observation vectors, the **sample mean** of the observation vectors is the vector:

$$M := \frac{1}{N} \sum_{j=1}^{N} X_j$$

The matrix:

$$B := \begin{pmatrix} X_1 - M & X_2 - M & \dots & X_N - M \end{pmatrix}$$

is called the **mean-deviation form** of the matrix X of observations. The columns of the matrix B are often denoted by $\widehat{X}_j$.

Remark 1.2. When we think of a matrix of observations X, one can think of the columns X_j as one set of observations of m variables. Thus, the rows correspond to variables and the row X^i of the matrix are N observations of the i^{th} variable. In these applications, N is usually large.

Exercise 1. Given an $m \times N$ matrix X of observations, show that its mean deviation form has zero sample mean.

Definition 1.3. Given a vector of observations $x = (x_1, ..., x_N)$, let $\widehat{x}$ be the average of the observations. The **sample variance** of the observations is the quantity:

$$\mathrm{Var}(x) := \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \widehat{x})^2$$

Definition 1.4. Given two vectors of observations $x = (x_1, ..., x_N)$ and $y = (y_1, ..., y_K)$ with means $\widehat{x}$ and $\widehat{y}$ respectively. The (sample) **covariance** of the observations is the quantity:

$$\mathrm{Covar}(x, y) := \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \widehat{x})(y_i - \widehat{y})$$

When the covariance between x and y is 0 we say the data x and y are **uncorrelated**.

Definition 1.5. Given an $m \times N$ matrix of observations X, let B be the mean-deviation form of X. The **sample-covariance matrix** of the matrix of observations is the $m \times m$matrix S defined by:

$$S := \frac{1}{N-1} BB^T$$

Exercise 2. Show that the sample covariance matrix of a matrix of observations is positive semidefinite (Use exercise 24 or 31 from Day 2.)

Exercise 3. Let X be a matrix of observations and let S be the covariance matrix of X. Suppose that the matrix X is already in mean-deivation form. Show that the diagonal entry S_{ii} of the matrix S corresponds to the variance of the i^{th} row of X viewed as a vector of observations. Show that the off-diagonal entry S_{ij} of the covariance matrix corresponds to the covariance of the i^{th} and j^{th} row of X.

Definition 1.6. Let X be an $m \times N$ matrix of observations and S its covariance matrix. The **total variance** of the observation matrix X is the trace:

$$\operatorname{tr}(S) := \sum_{j=1}^{m} S_{jj}$$

Exercise 4. Let A and B be two $n \times n$ matrices.

1. Show that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

2. Show that if A and B are similar, $\operatorname{tr}(A) = \operatorname{tr}(B)$.

Exercise 5. Let X be an $m \times N$ matrix of observations already in mean-deviation form and let P be an $m \times m$ orthogonal matrix. Let Y be the $m \times N$ matrix $Y := P^T X$. Then:

1. Show that the matrix Y is in mean-devation form.

2. Show that if the covariance matrix of X is S, then the covariance matrix of Y is $P^T S P$.

3. Show that the total variances of X and Y are the same.

Definition 1.7. Let X be an $m \times N$ matrix of observations and let S be its covariance matrix. Let S have the eigenvalues:

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m \geq 0$$

with corresponding orthonormal eigenvectors $u_1, u_2, ..., u_m$. The eigenvector u_i is called the i^{th} **principal component** of the data.

Remark 1.8. Principal component analysis consists of taking a matrix of observations X and finding an orthogonal change of variables $Y = P^T X$ that makes the new variables uncorrelated. The reason for requiring orthogonality can be seen in 5. We also want to put them in order of decreasing variance for the sake of choosing a convention.

Exercise 6. Let X be a matrix of observations in mean-deviation form and let S be its covariance matrix. Use facts about symmetric matrices from Day 2, to show there exists an orthogonal change of variables $Y = P^T X$ such that the matrix P consists of the principal components and the new covariance matrix shows the new variables are uncorrelated.

Exercise 7. Consider the following matrix of observations:

$$X = \begin{pmatrix} 19 & 22 & 6 & 3 & 2 & 20 \\ 12 & 6 & 9 & 15 & 13 & 5 \end{pmatrix}$$

1. Convert the matrix of observations to mean-deviation form.

2. Construct the sample covariance matrix.

3. Find the principal components of the data.

4. Perform a change of variables to principal components.

Remark 1.9. Given a matrix of observations, dimensional reduction consists of performing a change of variables to principal components and then orthogonally projecting to the subspace with the overlwhelming amount of variance.

Exercise 8. Suppose a 3×1000 matrix of observations X has the following covariance matrix:

$$S = \begin{pmatrix} 70 & 0 & 0 \\ 0 & 20 & 5\sqrt{3} \\ 0 & 5\sqrt{3} & 10 \end{pmatrix}$$

1. Obtain the principal components.

2. In the new variables, what are the proportions of each of the variances to the total variance?

3. Should we do dimensional reduction? To which subspace should we project?

4. Starting from the matrix of observations X, what does it mean to reduce dimensions as in remark 1.9? That is, what observations do we consider when we perform dimensional reduction on X?

Exercise 9. Let X be an $m \times N$ matrix of observations. Let $A := \frac{1}{\sqrt{N-1}} X^T$. Suppose $A = U\Sigma V^T$ is a singular value decomposition of A. Identify the eigenvalues of the covariance matrix and the principal components from the singular value decomposition.

Remark 1.10. Computing an SVD is better numerically than computing the eigenvalues of S. Thus, reading off the principal components from an SVD, as in the previous exercise, is often done in practice.

2 A brief look at Markov chains

Definition 2.1. Given vectors $v_1, ..., v_n \in \mathbb{R}^n$. A **convex combination** is a vector:

$$c_1 v_1 + ... + c_n v_n$$

such that the scalars c_i are nonnegative and sum to 1. The **convex hull** of the vectors $B = \{v_1, ..., v_n\}$ is denoted by $\text{Conv}(B)$ and consists of all convex combinations of

Exercise 10. Let A and B be sets in $\mathbb{R}^n$.

1. Show that a set B is convex if and only if $B = \text{conv}(B)$.

2. If $A \subseteq B$ and B is convex then $\text{conv}(A) \subseteq B$. Find a counterexample in $\mathbb{R}^2$ to show that equality need not hold.

3. If $A \subseteq B$ then $\text{conv}(A) \subseteq \text{conv}(B)$

4. $\text{conv}(A) \cup \text{conv}(B) \subseteq \text{conv}(A \cup B)$

Exercise 11. Consider the standard basis vectors $B = \{e_1, ..., e_n\}$ of $\mathbb{R}^n$. Describe geometrically, the convex hull $\text{Conv}(B)$. This is often denoted by Δ^{n-1}.

Definition 2.2. A vector $x \in \mathbb{R}^n$ is a **probability vector** or **stochastic vector** if it consists of nonnegative entries that add up to 1. That is, a stochastic vector is a vector in the convex hull of the standard basis vectors $e_1, ..., e_n$. We'll denote by $\mathcal{M}^n$ the set of stochastic vectors (that is, the standard simplex Δ^{n-1} in $\mathbb{R}^n$). A **stochastic matrix** is an $n \times n$ matrix P with columns that are stochastic vectors.

Exercise 12. Let P be an $n \times n$ stochastic matrix and let S be the $1 \times n$ matrix:

$$S = \begin{pmatrix} 1 & 1 & ... & 1 \end{pmatrix}$$

1. Show that $x \in \mathbb{R}^n$ is a stochastic vector if and only if $Sx = 1$ and all entries are nonnegative.
2. Show that $SP = S$.

Exercise 13. Let P be a stochastic matrix. Then:

1. Show that all nonnegative powers of P are also stochastic.
2. Show that if x is a stochastic vector, then so is Px.
3. Show that the product of stochastic matrices is a stochastic matrix.

Exercise 14. Let P be an $n \times n$ stochastic matrix. Show that 1 is an eigenvalue of P. (Hint: Instead of P consider P^T and the vector with every entry equal to 1.)

Exercise 15. Give an example of a stochastic matrix where the algebraic multiplicity of the eigenvalue 1 is greater than 1.

Exercise 16. Let P be an $n \times n$ stochastic matrix. Show that the eigenvalues λ of P are such that $|\lambda| \leq 1$. (Hint: Try a proof by contradiction.)

Exercise 17. Give an example of a stochastic matrix that has negative eigenvalues.

Exercise 18. Give an example of a stochastic matrix that is not invertible. Thus, a stochastic matrix may have eigenvalues equal to 0.

Exercise 19. Give an example of a stochastic matrix that isn't diagonalizable. (Hint: Look for a 3×3 stochastic matrix whose row reduced form for which you can easily determine the geometric and algebraic multiplicities.)

Definition 2.3. A **Markov chain** is a sequence of stochastic vectors $\{x_k\}_{k=0}^{\infty}$ and a stochastic matrix P such that:

$$x_{k+1} = Px_k$$

for all $k \geq 0$. Equivalently, a Markov chain is a pair (P, x_0) consisting of an $n \times n$ stochastic matrix P and a stochastic vector $x_0 \in \mathcal{M}^n$.

Remark 2.4. The entries in a stochastic vector are often viewed as possibles states of a system. Thus, it is often called a **state vector**. In a Markov chain, the sequence of state vectors exhibits how the probabilities of being in each of the states is changing over time. This is essentially a discrete dynamical system with initial condition x_0 and map the stochastic matrix P. Thus, it is natural to be interested in its long-term behavior. We think of the entry of P in position (i, j) as the probability to transition from the state corresponding to position j to the state corresponding to position i.

Definition 2.5. Given a Markov chain (P, x_0), a **steady state vector** is a fixed point of the matrix P. That is, a steady state is a vector $q \in \mathcal{M}^n$ such that $Pq = q$.

Exercise 20. Find the steady state vector of the stochastic matrix:

$$P = \begin{pmatrix} .7 & .1 & .3 \\ .2 & .8 & .3 \\ .1 & .1 & .4 \end{pmatrix}$$

Exercise 21. Show that every 2×2 stochastic matrix has a steady state vector.

Definition 2.6. Given a Markov chain (P, x_0) and a stochastic vector $q \in \mathcal{M}^n$. We say the Markov chain **converges** to the vector q if:

$$\lim_{k \to \infty} \left|\left|P^k x_0 - q\right|\right| = 0$$

That is, if the sequence $\{P^k x_0\}_{k=1}^{\infty}$ converges to q in the usual sense.

Definition 2.7. Let q be a steady state vector of a stochastic matrix P such that for an initial condition $x_0 \in \mathcal{M}^n$, the Markov chain (P, x_0) converges to q. The entries of q are called **long run probabilities** of the Markov chain.

Definition 2.8. A stochastic matrix P is **regular** if there exists an integer $k \geq 1$ such that P^k has only positive entries.

Exercise 22. Give an example of a stochastic matrix that is regular but not invertible.

Theorem 2.9. Let P be an $n \times n$ regular stochastic matrix, then there exists a steady state vector q of P. Furthermore, for any stochastic vector x_0, the Markov chain (P, x_0) converges to the steady state q.

Exercise 23. Consider the Markov chain given by the stochastic matrix P of exercise 20 and the initial vector $x_0 = \begin{pmatrix} 0.1 & 0.8 & 0.1 \end{pmatrix}^T$ converges to the steady statevector you found.

Theorem 2.10. Let P be an $n \times n$ stochastic matrix. Then P has a steady state stochastic vector $q \in \mathcal{M}^n$.

Exercise 24. Prove Theorem 2.10 as follows:

1. Let P be a stochastic matrix. Let $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$ be any vector and let $Px = (y_1, ..., y_n)$. Show that:
$$|y_1| + ...|y_n| \leq |x_1| + ... + |x_n|$$
with equality if and only if all the nonzero entries of x have the same sign.

2. Let v be an eigenvector for the eigenvalue 1. Apply part (1) and construct the steady state stochastic vector from v.

3 LU Factorizations

Definition 3.1. A matrix A is in **echelon form** if it has the following three properties:

1. Any nonzero rows are above zero rows.
2. A row's leading entry is in a column to the right of the leading entry of the row above it.
3. All entries below a leading entry are zero.

Definition 3.2. Let A be an $m \times n$ matrix. An **LU factorization** is a factorization of the form $A = LU$ where L is an $m \times m$ lower triangular matrix with ones on the diagonal and U is an $m \times n$ echelon form of A.

Example 3.3. The following is an LU decomposition:

$$\begin{pmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 0 \\ -3 & 8 & 3 & 1 \end{pmatrix} \begin{pmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & -1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Definition 3.4. An $n \times n$ **elementary matrix** is a matrix obtained by doing only one of the following to the $n \times n$ identity matrix:

1. Adding a multiple of a row to another row.
2. Interchanging two rows.
3. Multiplying all entries in a row by a nonzero constant.

Exercise 25. Let E be an $m \times m$ elementary matrix and A be an $m \times n$ matrix. What is the relationship of EA to A? What can you guarantee if E is lower triangular?

Theorem 3.5. Suppose A is an $m \times n$ matrix for which there exist unit lower-triangular elementary matrices $E_1, ...E_p$ such that:

$$E_p...E_1A$$

is an echelon form of A. Then $U := E_p...E_1A$ and $L := (E_p...E_1)^{-1}$ is an LU factorization of A.

Exercise 26. By row-reducing the matrix A use theorem 3.5 to find an LU factorization of the following:

1. $A = \begin{pmatrix} 3 & -7 & -2 \\ -3 & 5 & 1 \\ 6 & -4 & 0 \end{pmatrix}$
2. $A = \begin{pmatrix} 2 & -6 & 4 \\ -4 & 8 & 0 \\ 0 & -4 & 6 \end{pmatrix}$

Exercise 27. Suppose A is an $m \times n$ matrix and $b \in \mathbb{R}^m$ is a vector for which we want to solve the equation $Ax = b$. Suppose $A = LU$ is an LU factorization of A.

1. Turn the equation $Ax = b$ into a pair of equations one that involves L only and one that involves U only.

2. Give an algorithm to solve the system using your answer in (1).

3. Why do you think this is a good approach? Lay has a good brief discussion of the utility of the LU factorization [L94, Sec. 2.5].

Exercise 28. Use the method of exercise 27 to solve the system $Ax = b$ for each of the matrices there, where b is respectively:

$$b = \begin{pmatrix} -7 \\ 5 \\ 2 \end{pmatrix} \qquad b = \begin{pmatrix} 2 \\ -4 \\ 6 \end{pmatrix}$$

4 Duals and annihilators

Definition 4.1. Let V and W be vector spaces. The vector space of linear maps from V to W is the space:

$$\mathrm{Hom}(V, W) := \{T : V \to W \mid T \text{ is linear}\}$$

Exercise 29. Show that if V and W are vector spaces, then $\mathrm{Hom}(V, W)$ really is a vector space.

Exercise 30. Let V and W be finite-dimensional vector spaces. Show that $\mathrm{Hom}(V, W)$ is isomorphic (as vector spaces) to the space of matrices $\mathrm{Mat}(\dim W, \dim V)$ of $\dim W \times \dim V$ matrices. (Hint: Pick bases).

Exercise 31. Let V and W be finite-dimensional vector spaces. What is the dimension of $\mathrm{Hom}(V, W)$?

Definition 4.2. Let V be a finite-dimensional vector space. The dual of V is the vector space:

$$V^* := \mathrm{Hom}(V, \mathbb{R}) = \{T : V \to \mathbb{R} \mid T \text{ is linear}\}$$

Exercise 32. Let $\{e_1, ..., e_n\}$ be a basis of a finite-dimensional vector space V. Show there exists a basis $\{e_1^*, ..., e_n^*\}$ of the dual V^* such that:

$$e_i^*(e_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Conclude that V and V^* are isomorphic as vector spaces and that $\dim V = \dim V^*$.

Definition 4.3. For a finite-dimensional vector space V with a basis $\{e_i\}$. The basis $\{e_i^*\}$ of exercise 32 is called the **dual basis** of V dual to $\{e_i\}$.

Exercise 33. What is the dual basis in the following cases:

1. $\mathbb{R}^n$ with the standard basis vectors $\{e_1, ..., e_n\}$.

2. $\mathbb{R}[n]$ with the basis $\{1, t, t^2, ..., t^n\}$

Exercise 34. Let V and W be finite-dimensional vector spaces and let $T : V \to W$ be a linear map.

1. Show that for any linear map $\rho : W \to \mathbb{R}$, we get a linear map:

$$T^*\rho : V \to \mathbb{R}, \qquad T^*\rho(v) := \rho(T(v))$$

2. Show that the map:

$$T^* : W^* \to V^*, \qquad \rho \mapsto T^*\rho$$

is a linear map.

Exercise 35. Show that if $T : U \to V$ and $S : V \to W$ are linear maps between finite-dimensional vector spaces then $(S \circ T)^* = T^* \circ S^*$

Exercise 36. Let V be a finite-dimensional vector space. Consider the double dual $V^{**} = \mathrm{Hom}(\mathrm{Hom}(V, \mathbb{R}), \mathbb{R})$.

1. Let $v \in V$ be a fixed vector. Prove that the evaluation map:

$$\mathrm{ev}_v : V^* \to \mathbb{R}, \qquad \mathrm{ev}_v(\rho) := \rho(v)$$

is an element of V^{**}. That is, prove that ev_v is a linear map from V^* to $\mathbb{R}$.

2. Prove that the map:

$$N : V \to V^{**}, \qquad N(v) := \mathrm{ev}_v$$

is a linear isomorphism. (In fact, this map exhibits a natural isomorphism.)

Definition 4.4. Let V be a vector space and let $U \subseteq V$ be a subspace. The **annihilator** of U is the subspace:

$$U^0 := \{\rho \in V^* \mid \rho|_U \equiv 0\}$$

Exercise 37. Let V be a vector space with a subspace $U \subseteq V$. Show that the annihilator U^0 really is a vector space.

Exercise 38. Let V be a vector space. What are the annihilators of $\{0\}$ and V respectively?

Exercise 39. Let V be a finite-dimensional vector space. Suppose $U \subseteq V$ is a subspace different from $\{0\}$. Show that $U^0 \neq V^*$.

Exercise 40. Let V be a finite-dimensional vector space and let $U \subseteq V$ be a subspace. Show that $V = U \oplus U^0$.

Exercise 41. Let V be a finite-dimensional vector space and let U and W be subspaces of V. Show that if $U \subseteq W$, then $W^0 \subseteq U^0$.

5 Some multilinear algebra

Definition 5.1. Let $V_1, ..., V_n$ and U be vector spaces. A map:

$$T : V_1 \times ... \times V_n \to U$$

is **multilinear** (we sometimes say n**-linear**) if for all fixed $n-1$ vectors $v_1, ..., v_{i-1}, v_{i+1}, ..., v_n \in V$, the map:

$$V_i \to U, \qquad w \mapsto T(v_1, ..., v_{i-1}, w, v_{i+1}, ..., v_n)$$

is linear. That is, if freezing all but one variable yields a linear function.

Exercise 42. Show that the dot product on $\mathbb{R}^n$ is a bilinear (that is, a 2-linear) map.

Exercise 43. Make the identification:

$$\mathbb{R}^{n^2} \cong \underbrace{\mathbb{R}^n \times ... \times \mathbb{R}^n}_{n \text{ times}}$$

So that an element of $\mathbb{R}^{n^2}$ is viewed as n column vectors. Show that the determinant:

$$\det : \mathbb{R}^{n^2} \to \mathbb{R}, \qquad (v_1, ..., v_n) \mapsto \det\big(v_1|...|v_n\big)$$

is an n-linear map. The space of n-linear maps will be denoted by $\mathrm{Mult}(V_1 \times ... \times V_n, U)$.

Exercise 44. Let V, W, U be finite-dimensional vector spaces with bases $\{v_i\}_{i=1}^m$, $\{w_j\}_{j=1}^n$, $\{u_k\}_{k=1}^l$ and corresponding dual bases $\{v_i^*\}_{i=1}^m$, $\{w_j^*\}_{j=1}^n$, $\{u_k^*\}_{k=1}^l$. Show that the maps:

$$\phi_{i,j}^k : V \times W \to U, \qquad \phi_{i,j}^k(v, w) := v_i^*(v) w_j^*(w) u_k$$

are a basis of $\mathrm{Mult}(V \times W, U)$. Conclude that $\dim \mathrm{Mult}(V \times W, U) = \dim V \dim W \dim U$.

Definition 5.2. Let $V_1, ..., V_n$ and U be vector spaces and $T : V_1 \times ... \times V_n \to U$ be a multilinear map. The map T is **alternating** if whenever two of its arguments are permuted, the value switches sign. That is, for all $v_1, ...v_i, ...v_j, ..., v_n$ we have:

$$T(v_1, ..., v_i, ..., v_j, ..., v_n) = -T(v_1, ..., v_j, ..., v_i, ..., v_n)$$

Exercise 45. Show that the determinant is an alternating n-linear map when viewed as in exercise 43

Exercise 46. Let $\omega : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear map. Show there exists a matrix B such that for all $u, v \in V$ we have:

$$\omega(u, v) = v^T B u$$

References

[Hal58] P.R. Halmos. Finite-Dimensional Vector Sapces. Reprinting of the 1958 second edition. *Undergraduate Texts in Mathematics*. Springer-Verlag, New York-Heidelberg, 1974.

[L94] D.C. Lay. Linear Algebra and its Applications. Fourth Edition. Addison-Wesley, 2012.

[Rom08] S. Roman. Advanced Linear Algebra. Third Edition. *Graduate Texts in Mathematics, 135.* Springer, New York, 2008.