Eigenvalues and Eigenvectors - SVD - PCA

Chapter 6, Linear Algebra for Everyone (G. Strang)
Chapter 8, Applied Linear Algebra (P.J. Olver and C. Shakiban)
Chapters 6 and 12, An Introduction to Statistical Learning, G. James et al.
Chapters 3 and 14, The Elements of Statistical Learning, T. Hastie et al.

October 30, 2024

Solution of $Av = \lambda v$

Given a **square matrix** A, we have these definitions and remarks.

- A scalar λ is called **eigenvalue** of A if there is a vector $v \neq \bar{0}$ with
$$Av = \lambda v,$$
the vector v is called **eigenvector** associated to λ.
- That means that the "transformation" of v under A is onto the line spanned by v.
- Eigenvectors associated to $\lambda \neq 0$ are in $C(A)$,
$$C(A) = \text{L.C. of columns} = \{Ab,\ b \in \mathbb{R}^n\}.$$

Singular Value Decomposition

Any m by n matrix A can be factored into

$$\boxed{A = U\Sigma V^t = \text{Orthogonal}_{m,m}\text{Diagonal}_{m,n}\text{Orthogonal}_{n,n}}$$

- $A^tA = V\Sigma^2V^t$; V is the eigenvector matrix for A^tA.
- $AA^t = U\Sigma^2U^t$; U is the eigenvector matrix for AA^t.
- Σ: diagonal matrix defined by the singular values, which are the square roots of the eigenvalues of A^tA.

Principal Components Analysis - Statistical concepts

Suppose that $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ represents a collection of n measurements of a single variable

Variance-Varianza

$$\sigma_x^2 := \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \frac{\|\mathbf{x} - \bar{\mathbf{x}}\|^2}{n-1}$$

Standard deviation-Desviación Típica

$$\sigma_x := +\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The smaller the variance or standard deviation, the less spread out the measurements, and hence the more accurately the mean is expected to approximate the true value of the physical quantity.

PCA

Let $\mathcal{X} = \left[\frac{X_1 - \bar{X}_1}{sd(X_1)}, \ldots, \frac{X_p - \bar{X}_p}{sd(X_p)}\right]_{n,p}$ be a data matrix standardizing the p different variables/predictors of n observations.

For example, we have 6 rows, each one represents a person, and column 1 is the age, column 2 the weight and column 3 the height. Another example, in image analysis, each row of the data matrix represents an individual image whose components are, say, gray scale data for the individual pixels.

Suppose that the variables are correlated. The main goal of the PCA is to find an orthogonal matrix V that determines a change of variable, and the new variables $Z_1, \ldots, Z_P$ are not correlated and arranged in order in decreasing variance. Thus every row of $\mathcal{X}$ (observation vector) is replaced by another $z_{i1}, \ldots, z_{ip}$ s.t.

$$\mathcal{X} V = \mathcal{Z}, \quad \mathcal{X} v_i = Z_i, \quad [X_{i1}, \cdots, X_{ip}]\, V = [z_{i1}, \cdots, z_{ip}]\,.$$

The new variables are called principal components of $\mathcal{X}$. The $z_{i,j}$ are called principal component scores.

How to compute $\mathcal{Z}$ and V?

- V: eigenvectors of the correlation matrix $\frac{1}{n-1}\mathcal{X}^t\mathcal{X}$. Also called rotation matrix.
- Variances: eigenvalues of $\frac{1}{n-1}\mathcal{X}^t\mathcal{X}$
- $\mathcal{Z}=\mathcal{X}V$.

Algorithmically, by computing the SVD of $\mathcal{X}$, $\mathcal{X} = U\Sigma V^t$, such that

$$\mathcal{X}V = U\Sigma = \mathcal{Z}:$$

- Columns of $\mathcal{Z}$: principal components.
- Columns of V: principal directions
- The variances of the p.c. $\mathcal{Z}$: the squares of the singular values times $\frac{1}{n-1}$,

$$\frac{1}{n-1}\mathcal{Z}^t\mathcal{Z} = \frac{1}{n-1}\Sigma^t U^t U\Sigma = \frac{1}{n-1}\Sigma^2.$$