Homework 2

DUE: SATURDAY, FEBRUARY 1, 11:59PM

For all the problems below we assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Department of Mathematical Sciences, Carnegie Mellon University
E-mail address: jkoganem@andrew.cmu.edu.
Date: January 26, 2025.

Problem 2.1 (Decorrelation of random variables and Gram-Schmidt)**.** Recall that if $\boldsymbol{v}_1, \boldsymbol{v}_2$ are vectors in $\mathbb{R}^n$, then the projection of $\boldsymbol{v}_2$ onto $\boldsymbol{v}_1$ is given by

$$\operatorname{proj}_{\boldsymbol{v}_1} \boldsymbol{v}_2 = \frac{\boldsymbol{v}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1. \tag{2.1}$$

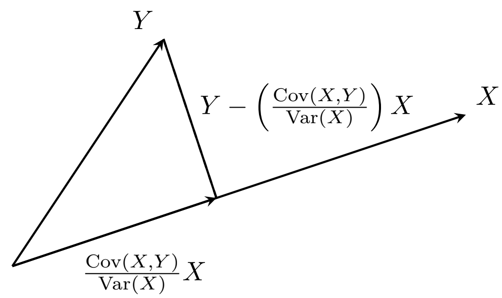
By construction, $\boldsymbol{v}_2 - \operatorname{proj}_{\boldsymbol{v}_1} \boldsymbol{v}_2$ is orthogonal to $\boldsymbol{v}_1$. This is the idea behind the Gram-Schmidt process. In this problem we explore the connection between the Gram-Schmidt process and the decorrelation of random variables.

Let $X, Y : \Omega \to \mathbb{R}$ be random variables with zero mean and finite variance. Define the random variable $Z : \Omega \to \mathbb{R}$ via

$$Z = Y - \left(\frac{\operatorname{Cov}(X, Y)}{\operatorname{Var}(X)} \right) X. \tag{2.2}$$

Show that $\operatorname{Cov}(X, Z) = 0$.

The following toy picture might be instructive:



You may also use the fact that expectation is linear: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ for any scalars $a, b \in \mathbb{R}$ and random variables $X, Y : \Omega \to \mathbb{R}$.

Problem 2.2 (Decorrelation and eigendecomposition of the covariance matrix)**.** Here's another way to think about decorrelation. Let $X_1, X_2 : \Omega \to \mathbb{R}$ be random variables with zero mean and finite variance. Consider the random vector $\boldsymbol{X} : \Omega \to \mathbb{R}^2$ defined by

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}. \tag{2.3}$$

The covariance matrix $\Sigma \in \mathcal{M}_{2\times 2}(\mathbb{R})$ of $\boldsymbol{X}$ is given by

$$\Sigma = \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}[X_2] \end{pmatrix} = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]. \tag{2.4}$$

a) Suppose Σ admits zero as an eigenvalue, which implies that $\det \Sigma = 0$. Show that $\mathrm{Corr}(X_1, X_2) = \pm 1$, i.e. X_1 and X_2 are perfectly correlated.

b) Now assume Σ is positive definite. By the spectral theorem, there exists an orthogonal matrix $Q \in \mathcal{M}_{2\times 2}(\mathbb{R})$ and an invertible diagonal matrix $\Lambda \in \mathcal{M}_{2\times 2}(\mathbb{R})$ such that $\Sigma = Q\Lambda Q^T$. Consider the random vector $\boldsymbol{Y} : \Omega \to \mathbb{R}$ defined via $\boldsymbol{Y} = Q^T\boldsymbol{X}$. Show that $\mathrm{Cov}(\boldsymbol{Y}) = \Lambda$ and conclude that $\mathrm{Cov}((\boldsymbol{Y})_1, (\boldsymbol{Y})_2) = 0$.

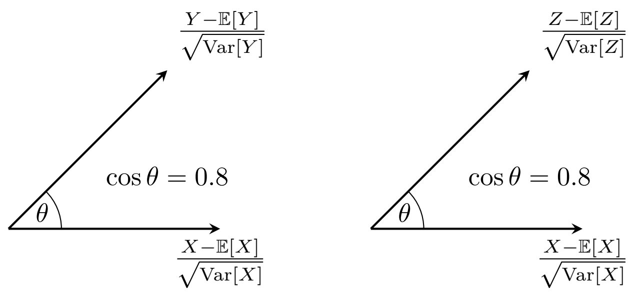
Note that part b) can easily be generalized to the case of n random variables. Also note that when one performs PCA, one is effectively performing eigen-decompositon on the sample covariance matrix. So from this point of view, PCA can also be thought of as a method to decorrelate the random variables from which the samples are drawn from.

Problem 2.3 (Maximum and minimum correlation)**.** Suppose $X, Y, Z : \Omega \to \mathbb{R}$ are random variables with finite mean and variance, $\mathrm{Corr}(X, Y) = 0.8$, and $\mathrm{Corr}(X, Z) = 0.8$. Let $\rho = \mathrm{Corr}(Y, Z)$. In this problem we want to find the maximum and minimum possible values of ρ.

From Problem 2.1, one can see that

a) $\mathrm{Cov}(X, Y)$ is analogous to the inner product of two vectors in $\mathbb{R}^n$; in particular, zero covariance corresponds to orthogonality
b) $\sqrt{\mathrm{Var}(X)}$ is analogous to the length of a vector in $\mathbb{R}^n$ (the square root of the variance is referred to as the standard deviation)
c) $\mathrm{Corr}(X, Y)$ is analogous to the cosine of the angle between two vectors in $\mathbb{R}^n$

Note of precaution: by definition, the covariance, variance, and correlation of two random variables are always measured relative to their respective means, so one should keep in mind that this analogy should only be made with respect to the *centered* random variables $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$. From the point of view of linear algebra, by centering we are effectively making sure that the elements of the underlying vector space are all centered at a common origin. For this reason we can draw the following toy pictures to help us solve this problem:

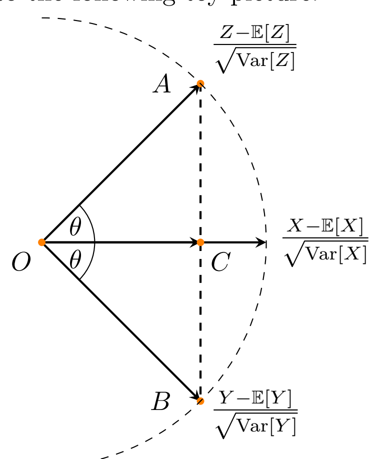


Note that

$$\mathrm{Var}\left[\frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]}}\right] = \left(\frac{1}{\sqrt{\mathrm{Var}[X]}}\right)^2 \mathrm{Var}[X - \mathbb{E}[X]] = \frac{1}{\mathrm{Var}[X]}\mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2]$$
$$= \frac{1}{\mathrm{Var}[X]}\left(\mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2\right) = \frac{1}{\mathrm{Var}[X]}\left(\mathbb{E}[X^2] - (\mathbb{E}[X])^2\right) = \frac{\mathrm{Var}[X]}{\mathrm{Var}[X]} = 1, \quad (2.5)$$

so $\frac{X-\mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]}}$ can be thought of as corresponding to a unit vector in $\mathbb{R}^n$.

a) Convince yourself that if we want to minimize $\mathrm{Corr}(Y, Z)$, we want to maximize the "angle" between $Y - \mathbb{E}[Y]$ and $Z - \mathbb{E}[Z]$. This leads to the following toy picture:



Our goal is to find $\cos 2\theta$. Note that the lengths $|OA| = |OB| = 1$. Find $|OC|, |AC|$ and use the identity $\cos 2\theta = (\cos\theta)^2 - (\sin\theta)^2$ to find the maximum value of $\mathrm{Corr}(Y, Z)$.

b) Convince yourself that if we want to maximize $\mathrm{Corr}(Y, Z)$, we want to minimize the "angle" between $Y - \mathbb{E}[Y]$ and $Z - \mathbb{E}[Z]$. What would be a choice of Y, Z such that the "angle" between $Y - \mathbb{E}[Y]$ and $Z - \mathbb{E}[Z]$ is minimized? Use this to find the minimum value of $\mathrm{Corr}(Y, Z)$.

Note that this technique can be used even if $\mathrm{Corr}(X, Y) \neq \mathrm{Corr}(X, Z)$. One can draw the same geometric diagram as in part a) and use the formula $\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ for $\alpha, \beta \in \mathbb{R}$.

Problem 2.4 (The correlation matrix)**.** Here is an alternative way to solve the previous problem. Same setup as before: suppose $X, Y, Z : \Omega \to \mathbb{R}$ are random variables with finite mean and variance, and suppose $\mathrm{Corr}(X, Y) = 0.8$ and $\mathrm{Corr}(X, Z) = 0.8$. Let $\rho = \mathrm{Corr}(Y, Z)$. In this problem we define the random vector $\boldsymbol{X} : \Omega \to \mathbb{R}^3$ via $\boldsymbol{X} = \begin{pmatrix} X & Y & Z \end{pmatrix}^T$.

a) Write down the correlation matrix $P \in \mathcal{M}_{3\times 3}(\mathbb{R})$ of $\boldsymbol{X}$.
b) Calculate $\det P$ in terms of ρ.
c) Use the fact that P is positive semi-definite (we will prove this in Problem 2.6) to conclude find the maximum and minimum possible values of ρ.

You should find the same maximum and minimum values of ρ as in Problem 2.3.

Problem 2.5 (Generating correlated random variables with eigen-decomposition)**.** In Problems 2.1 and 2.3 we explored how to decorrelate random variables. In this problem we explore the opposite problem: how to generate correlated random variables.

Fix $\boldsymbol{\mu} \in \mathbb{R}^n$ and a positive semi-definite matrix $\Sigma \in \mathcal{M}_{n\times n}(\mathbb{R})$. Suppose we want to generate a sequence of random variables $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ such that the random vector $\boldsymbol{X} = \begin{pmatrix} X_1 & \ldots & X_n \end{pmatrix}^T$ has mean $\boldsymbol{\mu}$ and covariance matrix Σ. We can achieve this by starting with a sequence of pairwise uncorrelated random variables $Z_1, Z_2, \ldots, Z_n : \Omega \to \mathbb{R}$ with mean 0 and variance 1 and then apply an affine transformation to obtain $X_1, X_2, \ldots, X_n$.

By the spectral theorem, there exists an orthogonal matrix $Q \in \mathcal{M}_{n\times n}(\mathbb{R})$ and an invertible diagonal matrix $\Lambda \in \mathcal{M}_{n\times n}(\mathbb{R})$ such that $\Sigma = Q\Lambda Q^T$. Define the random vector $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ via

$$\boldsymbol{X} = \boldsymbol{\mu} + Q\Lambda^{1/2}\boldsymbol{Z}, \tag{2.6}$$

where

$$\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}, \quad \Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\lambda_n} \end{pmatrix} \tag{2.7}$$

and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of Σ and the diagonal entries of Λ.

a) Show that $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu}$.
b) Show that $\mathrm{Cov}(\boldsymbol{X}) = \Sigma$.

This problem also shows that every positive semi-definite matrix is a covariance matrix, in the sense that one can construct a random vector $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ with covariance matrix Σ.

Problem 2.6. Let $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ be a random vector. The covariance matrix $\Sigma \in \mathcal{M}_{n\times n}(\mathbb{R})$ of $\boldsymbol{X}$ is defined via

$$\Sigma = \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T]. \tag{2.8}$$

In this problem we show that Σ is positive semi-definite.

a) Show that

$$\mathrm{Var}(\boldsymbol{X} \cdot \boldsymbol{x}) = \boldsymbol{x}^T \Sigma \boldsymbol{x}, \text{ for any } \boldsymbol{x} \in \mathbb{R}^n. \tag{2.9}$$

b) Use (2.9) to show that $\boldsymbol{x}^T \Sigma \boldsymbol{x} \geq 0$ for any $\boldsymbol{x} \in \mathbb{R}^n$, and conclude that Σ is positive semi-definite.

c) Let's explore what we can say about the random vector $\boldsymbol{X}$ if Σ is positive definite as opposed to being just positive semi-definite. Suppose Σ is not positive definite, i.e. there exists a non-zero vector $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{x}^T \Sigma \boldsymbol{x} = 0$. Use (2.9) and the analogy that $\mathrm{Var}(\cdot)$ is the square of the "length" of a random variable to conclude that $\boldsymbol{X} \cdot \boldsymbol{x} = 0$ (almost surely[1]).

d) Let

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \tag{2.10}$$

where $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ are random variables. Use part c) to conclude that Σ is positive definite iff. $X_1, \ldots, X_n$ are linearly independent as random variables (almost surely).

[1]This is a somewhat technical point, but in measure theoric probability, a random variable $Y : \Omega \to \mathbb{R}$ has "length" 0 iff. Y agrees with the zero random variable "almost everywhere", meaning that they differ only on a set of probability measure 0, which is negligible. In other words, $\mathbb{P}(\{\omega \in \Omega : Y(\omega) = 0\}) = \mathbb{P}(Y = 0) = 1$. Probabilists would say that $Y = 0$ *almost surely*. Since $\boldsymbol{X} \cdot \boldsymbol{x} : \Omega \to \mathbb{R}$ is a random variable, to be precise we can only conclude that $\boldsymbol{X} \cdot \boldsymbol{x} = 0$ almost surely.