ME599-004: Data-Driven Methods for Control Systems

Instructors: Uduak Inyang-Udoh, Yuxin Tong

January 8, 2025

Contents

2 Dimensionality Reduction

2.1 Principal Component Analysis (PCA)

2.1.1 Expectation

Expectation of a random variable is an average of all realizations of X weighted by their likelihoods:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, p_X(x)\, dx$$

Intuition:

Regular average: We sum all values and divide by the number of values.
Expectation: Weights values by their relative likelihoods and sums them up.
In the multivariate case:

$$\mathbb{E}[\mathbf{X}] = \left[\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_N]\right]$$

Sidenote: You can show that for a Gaussian distribution:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad dx = \mu$$

2.1.2 Variance

Measure of variability around the mean, weighted by likelihoods:

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2\, P_X(x) dx$$

In the multivariate case, this idea becomes the variance-covariance matrix:

$$\Sigma = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right]$$

We can think of the diagonal of Σ as the element-wise variance of the multivariate random variable:

$$\Sigma_{ii} = \mathrm{Var}[X(i)] = \mathbb{E}\left[(X(i))^2\right] - \mathbb{E}[X(i)]^2$$

i.e., a measure of variability of the ith entry.

2.1.3 Covariance

Covariance is the term used to express the relationship (with respect to the mean) between components of the vector random variable:

$$\Sigma_{i,j} = \mathbb{E}\left[(X(i) - \mathbb{E}[X(i)])\,(X(j) - \mathbb{E}[X(j)])\right]$$

If $\Sigma_{i,j} = 0$, the components are unrelated (orthogonal).
If $\Sigma_{i,j} > 0$ ($\Sigma_{i,j} < 0$), they are positively (negatively) related.
Note:

1. If $\mathbb{E}[X] = \vec{0}$, then the covariance matrix for any realization $\vec{x}$ is $\Sigma = \mathbb{E}[\vec{x}, \vec{x}^T]$.

2. Covariance matrix:

$$\Sigma = \mathbb{E}\left[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T\right], \quad \text{where } \vec{\mu} = \mathbb{E}[\vec{x}]$$

is symmetric, i.e., $\Sigma_{i,j} = \Sigma_{j,i}$, and is in fact positive semi-definite. (**Question:** Can you show this?)

2.1.4 Discrete Random Variables

Assuming there are a countable number N realizations of the random variable X, the mean is given by:

$$\mathbb{E}(X) = \sum_{i=1}^{N} P(X = \vec{x}_i)\vec{x}_i = \vec{\mu}$$

The covariance matrix is given by:

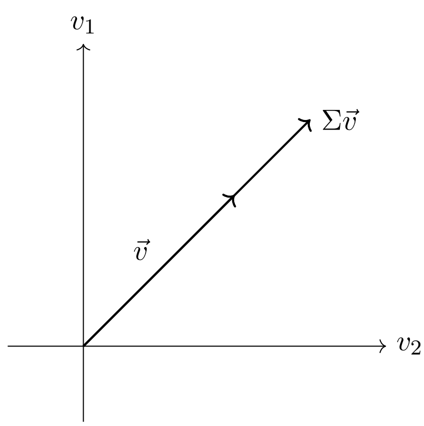
$$\Sigma = \sum_{i=1}^{N} P(X = \vec{x}_i)(\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T$$

(**Question:** Assuming $P(X = \vec{x}_i) = \frac{1}{N}$, $\forall i$, what are $\vec{\mu}$ and Σ?)

2.1.5 Eigenvalues & Eigenvectors of Covariance Matrix

Eigenvector: a vector $\vec{v}$ such that:

$$\Sigma\vec{v} = \lambda\vec{v}$$



where λ is the corresponding eigenvalue, $\Sigma\vec{v}$ and $\vec{v}$ are collinear.

If $\vec{v}$ is an eigenvector, $\alpha\vec{v}$ is also an eigenvector for any $\alpha \in \mathbb{R}$,

$$\Sigma(\alpha\vec{v}) = \alpha\Sigma\vec{v} = \alpha\lambda\vec{v} = \lambda(\alpha\vec{v}).$$

So we can typically normalize an eigenvector, without loss of generality, by dividing by its norm:

$$\frac{\vec{v}}{\|\vec{v}\|} \mapsto \vec{v} \quad \text{for } \|\vec{v}\|^2 \neq 1.$$

Note: 1. Because Σ is **PSD (positive semi-definite)** and symmetric, all its eigenvalues are real and nonnegative (check that this is so!). This is also the case for complex-valued random vectors (for which Σ is now Hermitian).

2. In addition, the eigenvectors are orthogonal.

Proof of Note 2: Assume $\vec{v}$ and $\vec{u}$ are eigenvectors associated with eigenvalues λ and μ, respectively, and $\lambda \neq \mu$, then

$$\Sigma\vec{v} = \lambda\vec{v}, \quad \Sigma\vec{u} = \mu\vec{u}.$$

Since the matrix Σ is Hermitian, we have that:

$$\vec{u}^\dagger\Sigma\vec{v} = \vec{u}^\dagger\Sigma^\dagger\vec{v} = (\Sigma\vec{u})^\dagger\vec{v}$$

Now, if we make $\Sigma\vec{v} = \lambda\vec{v}$ on the leftmost side, and $\Sigma\vec{u} = \mu\vec{u}$ on the rightmost side, we find that:

$$\vec{u}^\dagger\lambda\vec{v} = \lambda\vec{u}^\dagger\vec{v} = (\mu\vec{u})^\dagger\vec{v} = \mu\vec{u}^\dagger\vec{v}$$

This implies:

$$\lambda\vec{u}^\dagger\vec{v} = \mu\vec{u}^\dagger\vec{v}$$

The above equality holds only if:

$$\vec{u}^\dagger\vec{v} = 0,$$

since $\mu \neq \lambda$ and $\mu, \lambda \in \mathbb{R}$.

Thus, the eigenvectors $\vec{u}$ and $\vec{v}$ are orthogonal.

2.1.6 Eigenvector Matrix

Let T be the eigenvector matrix whose columns are the ordered eigenvectors of Σ. Since the eigenvectors are orthonormal:

$$T = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix},$$

and

$$T^\dagger T = I \iff T \text{ is unitary.}$$

2.1.7 PCA Transform

We define the **Principal Component Analysis Transform** as:

$$\vec{y} = T^\dagger\vec{x} \quad \Longleftarrow \quad \text{The projection onto the coordinates } v_0, \ldots, v_{n-1} \text{ (or 'basis').}$$

And the inverse, **iPCA Transform**:

$$\vec{\tilde{x}} = T\vec{y} \quad \Longleftarrow \quad \text{Scale the PCs (or 'basis') by } y_0, y_1, \ldots, y_{n-1}.$$

Since T is unitary:

$$\vec{\tilde{x}} = T\vec{y} = TT^\dagger\vec{x} = I\vec{x} = \vec{x}.$$

Thus, $\vec{y}$ is an equivalent representation of $\vec{x}$, and:

$$\|\vec{x}\|^2 = \|\vec{y}\|^2 \quad \Longleftarrow \quad \text{The energy composition of the vector is conserved.}$$

The PCA transform is defined for any vector $\vec{x}$ of appropriate dimension. But in particular, we expect it to give suitable principal components when $\vec{x}$ is a realization of X.

(**Question:** Do you recognize similarities with the DFT?)

$$\text{DFT:} \quad \hat{f}_k = \sum_{j=0}^{N-1} f_j \omega_N^{kj}, \quad \text{where } \omega_N = e^{-2\pi i/N},$$

$$\text{PCA:} \quad y_k = \sum_{j=0}^{n-1} x_j v_{k,j}.$$

Different 'bases' are more specific to a given random variable, but the sense of taking to a new 'basis' persists.

2.1.8 Compression with PCA

Recall that for the Discrete Fourier Transform (DFT), we compress by retaining $C < N$ DFT coefficients. When doing the inverse DFT (iDFT):

$$\tilde{f}_j = \sum_{k=0}^{C-1} \hat{f}_k \omega_N^{jk}.$$

In a similar fashion, we can compress using PCA by retaining $C < N$ PCA coefficients:

$$\tilde{\vec{x}}_j = \sum_{k=0}^{C-1} y_k v_{k,j},$$

or:

$$\tilde{\vec{x}} = T\tilde{\vec{y}},$$

where:

$$\tilde{y}_k = \begin{cases} y_k, & \text{for } k < C, \\ 0, & \text{otherwise.} \end{cases}$$

2.1.9 Review

In the last class, our discussion was as follows:

Supposing we have N realizations of random variable vectors $\vec{x} \in \mathbb{R}^n$ (random variable is a collection of all possible images of different kinds of logs in Michigan state), the mean or expected value based on the N realizations, or assuming only N realizations exist:

$$\mu = \sum_{i=1}^{N} P(X = \vec{x}_i)\vec{x}_i = \frac{1}{N}\sum_{i=1}^{N} \vec{x}_i.$$

The covariance matrix expresses the expectation of how any two components (or pixels) are related, i.e.,

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T.$$

We showed that the eigenvectors of Σ are orthonormal to one another because it is Positive Semi-Definite (PSD) and Hermitian; and the eigenvalues are real and non-negative.

We noted that these eigenvectors can serve as a new coordinate system or basis for our random variable space; and more importantly, the eigenvectors are ordered in terms of relative importance of how much they express distinction or variance in the data.

In this way, we order the eigenvectors according to corresponding eigenvalues.

Note: Unbiased covariance uses $\frac{1}{N-1}$.

$$T = \begin{bmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n-1}\end{bmatrix} \quad \vec{v}_1, \vec{v}_2 \text{ ... are Principal Components.}$$

The PCA transformation is given by:

$$\vec{y} = T^\dagger \vec{x} \quad \longleftarrow \quad \text{PCA transformation (projection onto the eigenvector space).}$$

Further:

$$\Sigma = T\Lambda T^\dagger,$$

where Λ is the diagonal matrix of eigenvalues.

2.1.10 Variance of Transformed Data

Notice that:

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T$$

is equivalent to:

$$\Sigma = \frac{1}{N}\left(X - \vec{\mu}\vec{1}^T\right)\left(X - \vec{\mu}\vec{1}^T\right)^T,$$

where:

$$X = \begin{bmatrix}\vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_N\end{bmatrix}, \quad \vec{1} = \underbrace{\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}^T}_{n}$$

Thus:

$$\Sigma = \frac{1}{N}X_m X_m^T,$$

where X_m is the mean-subtracted data.

In the PCA-transformed space, the data has zero covariance:

$$Y_m = T^\dagger X_m.$$

Then:

$$\Sigma_y = \frac{1}{n}Y_m Y_m^T = \frac{1}{n}{T_n}^\dagger X_m X_m^\dagger T_n = {T_n}^\dagger \Sigma_x T_n.$$

But recall that:

$$\Sigma_x = T_n \Lambda {T_n}^\dagger.$$

Therefore:

$$\Sigma_y = {T_n}^\dagger \Sigma_x T_n = {T_n}^\dagger T_n \Lambda {T_n}^\dagger T_n = \Lambda,$$

Since Σ_y is diagonal, it contains only variances (no covariances).

2.2 Singular Value Decomposition (SVD)

Every complex-valued matrix $A \in \mathbb{C}^{n\times m}$ can be decomposed as:

$$A = USV^\dagger,$$

where $U \in \mathbb{C}^{n\times n}$ and $V \in \mathbb{C}^{m\times m}$ are unitary matrices with orthonormal columns, and $S \in \mathbb{R}^{n\times m}$ is a matrix with real, non-negative entries on the diagonals and zeros off the diagonal.

When $n > m$, the matrix S has at most m non-zero elements on the diagonal and may be written as:

$$S = \begin{bmatrix} \hat{S} \\ 0 \end{bmatrix}.$$

Therefore, we can exactly represent the matrix A using the 'economy' SVD:

$$A = USV^\dagger = \begin{bmatrix} \hat{U} & \hat{U}^\perp \end{bmatrix} \begin{bmatrix} \hat{S} \\ 0 \end{bmatrix} V^\dagger = \hat{U}\hat{S}V^\dagger.$$

$\begin{bmatrix} \hat{U} & \hat{U}^\perp \end{bmatrix}$ has dimensions $n \times n$, $\begin{bmatrix} \hat{S} \\ 0 \end{bmatrix}$ has dimensions $n \times m$, $V^\dagger$ has dimensions $m \times m$.

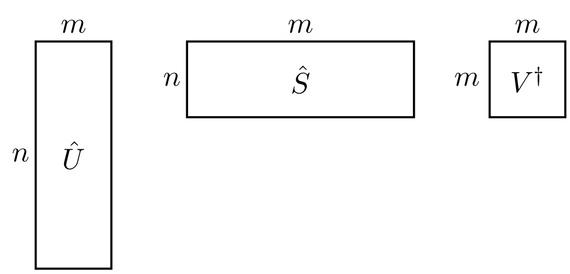
Thus:

$\hat{U} \in \mathbb{C}^{n\times m}$: Contains the first m left singular vectors of A, which span the column space of A.

$\hat{S} \in \mathbb{R}^{n\times m}$: A diagonal matrix containing the m singular values.

$V^\dagger \in \mathbb{C}^{m\times m}$: The conjugate transpose of V, which is a unitary matrix whose columns are the right singular vectors of A.

$\hat{U}^\perp$ represents the orthogonal complement of $\hat{U}$, the columns of $\hat{U}^\perp$ span a vector space that is complementary and orthogonal to that spanned by $\hat{U}$.



The columns of U are called *left singular vectors*; those of V are *right singular vectors*. The diagonal elements of $\hat{S}$ are called *singular values* and are ordered from largest to smallest. The rank of matrix A is equal to the number of non-zero singular values.

We can write the SVD for the mean-subtracted data as:

$$X_m = USV^\dagger$$

Recall that the covariance matrix is:

$$\Sigma = \frac{1}{n} X_m X_m^\dagger = \frac{1}{n} US^2U^\dagger = T\Lambda T^\dagger$$

Hence, the principal components of T are also the left singular vectors of X_m. The variance of the data in the new coordinates is:

$$\lambda_k = \frac{\sigma_k^2}{n}$$

References

[1] SL. Brunton , JN. Kutz *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press; 2019.

[2] A. Mokhtari, S. Paternain and A. Ribeiro, “Principal Component Analysis [lecture notes],” 2018; http://www.seas.upenn.edu/users/ aribeiro/.