Mathematics for Machine Learning

— Linear Algebra: Projections & Gram-Schmidt Orthogonalization

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,
Tamkang University

Fall 2023

Credits for the resource

- The slides are based on the textbooks:
  - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
  - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.*
- We could partially refer to the monograph:
  Francesco Orabona: A Modern Introduction to Online Learning.
  https://arxiv.org/abs/1912.13213

Outline

Outline

Motivations (1/2)

- In machine learning, we often need to deal with *high-dimensional* data.
- High-dimensional data is often hard to analyze or visualize.
- Sometimes, only a few dimensions contain most information.
- We might try to project the original high-dimensional data onto a lower-dimensional space and work on it.
- Note: When we compress or visualize high-dimensional data, we will lose information.

Motivations (2/2)

Examples (dimensionality reduction)

- Principal component analysis (PCA)
- Deep neural networks
- Classification
- Linear Regression

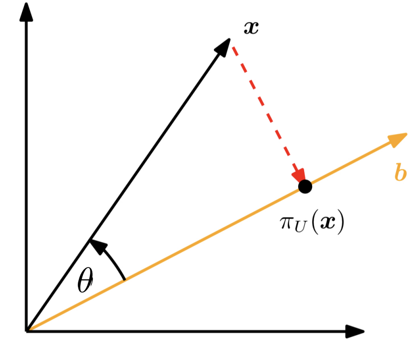
Projection

Projection

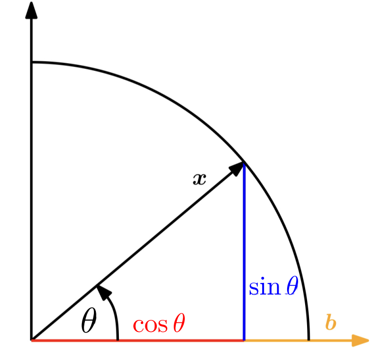
Let V be a vector space and $U \subseteq V$ be a subspace of V. A linear mapping $\pi : V \mapsto U$ is called a projection if $\pi^2 = \pi \circ \pi = \pi$.

- Recall that linear mappings can be expressed by transformation matrices.
- The projection matrices $\boldsymbol{P}_\pi$ exhibit the property that $\boldsymbol{P}_\pi^2 = \boldsymbol{P}_\pi$.

Illustration of projections onto 1-D



(a) Projection of $\boldsymbol{x} \in \mathbb{R}^2$ onto a subspace U with basis vector $\boldsymbol{b}$.



(b) Projection of a two-dimensional vector $\boldsymbol{x}$ with $\|\boldsymbol{x}\| = 1$ onto a one-dimensional subspace spanned by $\boldsymbol{b}$.

Illustration of projections onto 1-D

- $\pi_U(\mathbf{x})$: closest to $\mathbf{x}$.
  - $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ is minimal.
  - $\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0$.
- Projection $\pi_U(\mathbf{x})$ of $\mathbf{x}$ onto U must be an element in U.
  - $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$.
- Determining the coordinates:

  Since $\pi_U(\mathbf{b}) = \lambda \mathbf{b}$:

$$\langle \mathbf{x} - \pi_U(\mathbf{x}), \mathbf{b} \rangle = 0 \;\Leftrightarrow\; \langle \mathbf{x} - \lambda \mathbf{b}, \mathbf{b} \rangle = 0.$$

$$\Leftrightarrow\; \langle \mathbf{x}, \mathbf{b} \rangle - \lambda \langle \mathbf{b}, \mathbf{b} \rangle = 0 \;\Leftrightarrow\; \lambda = \frac{\langle \mathbf{x},\mathbf{b} \rangle}{\langle \mathbf{b},\mathbf{b} \rangle} = \frac{\langle \mathbf{b},\mathbf{x} \rangle}{\|\mathbf{b}\|^2} = \frac{\mathbf{b}^\top \mathbf{x}}{\mathbf{b}^\top \mathbf{b}}.$$

- Finding the projection $\pi_U(\mathbf{x}) \in U$:

$$\pi_U(\mathbf{x}) = \lambda \mathbf{b} = \frac{\langle \mathbf{x}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2} \mathbf{b} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} \mathbf{b},$$

  Note that $\|\pi_U(\mathbf{x})\| = \|\lambda \mathbf{b}\| = |\lambda| \|\mathbf{b}\|$.

- If we use the dot product as the inner product and let θ be the angle between $\mathbf{x}$ and $\mathbf{b}$:

$$\|\pi_U(\mathbf{x})\| = \frac{|\mathbf{b}^\top \mathbf{x}|}{\|\mathbf{b}\|^2} \|\mathbf{b}\| = |\cos\theta| \|\mathbf{x}\| \|\mathbf{b}\| \frac{\|\mathbf{b}\|}{\|\mathbf{b}\|^2} = |\cos\theta| \|\mathbf{x}\|.$$

- Finding the projection matrix $\boldsymbol{P}_\pi$:
  - Recall: projection is a linear mapping.
  - With the dot product as the inner product,

$$\|\pi_U(\mathbf{x})\| = \lambda \mathbf{b} = \mathbf{b}\lambda = \mathbf{b}\frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} = \frac{\mathbf{b}\mathbf{b}^\top}{\|\mathbf{b}\|^2}\mathbf{x}.$$

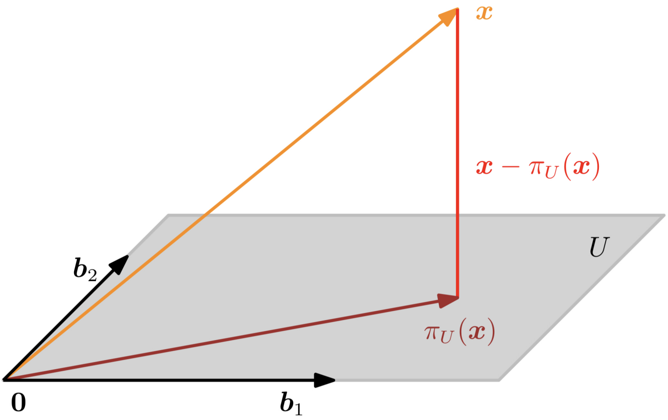
  - So,

$$\boldsymbol{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\|\mathbf{b}\|^2}.$$

Note: $\mathbf{b}\mathbf{b}^\top$ is a symmetric matrix.

Projection onto General Subspaces (1/4)

Orthogonal projections of $\mathbf{x} \in \mathbb{R}^n$ onto $U \subseteq \mathbb{R}^n$ with $\dim(U) = m \geq 1$.



Projection onto General Subspaces (2/4)

- Any projection can be represented as a linear combination of the basis vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m$ of U.
  - $\pi_U(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \mathbf{b}_i$.
- Find the coordinates $\lambda_1, \ldots, \lambda_m$:

$$\pi_U(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \mathbf{b}_i = \boldsymbol{B}\boldsymbol{\lambda} \ \text{(closest to } \mathbf{x} \text{ on } U\text{)}$$

for $\boldsymbol{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_m] \in \mathbb{R}^{n \times m}$, $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]^\top \in \mathbb{R}^m$.

Note: $\mathbf{x} \perp \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ ($\because$ minimum distance)

$$\begin{aligned} \langle \mathbf{b}_1, \mathbf{x} - \pi_U(\mathbf{x}) \rangle &= \mathbf{b}_1^\top (\mathbf{x} - \pi_U(\mathbf{x})) = 0 \\ &\vdots \\ \langle \mathbf{b}_m, \mathbf{x} - \pi_U(\mathbf{x}) \rangle &= \mathbf{b}_m^\top (\mathbf{x} - \pi_U(\mathbf{x})) = 0 \end{aligned}$$

Projection onto General Subspaces (3/4)

- Any projection can be represented as a linear combination of the basis vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m$ of U.
  - $\pi_U(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \mathbf{b}_i$.
- Find the coordinates $\lambda_1, \ldots, \lambda_m$:

$$\pi_U(\mathbf{x}) = \sum_{i=1}^{m} \lambda_i \mathbf{b}_i = \boldsymbol{B}\boldsymbol{\lambda} \quad \text{(closest to } \mathbf{x} \text{ on } U\text{)}$$

for $\boldsymbol{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_m] \in \mathbb{R}^{n \times m}$, $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]^\top \in \mathbb{R}^m$.

Note: $(\mathbf{x} - \pi_U(\mathbf{x})) \perp \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ ($\because$ minimum distance)

$$\begin{aligned} \mathbf{b}_1^\top(\mathbf{x} - \boldsymbol{B}\boldsymbol{\lambda}) &= 0 \\ &\vdots \\ \mathbf{b}_m^\top(\mathbf{x} - \boldsymbol{B}\boldsymbol{\lambda}) &= 0 \end{aligned}$$

Projection onto General Subspaces (4/4)

Since

$$\begin{aligned} \mathbf{b}_1^\top(\mathbf{x} - \boldsymbol{B\lambda}) &= 0 \\ &\vdots \\ \mathbf{b}_m^\top(\mathbf{x} - \boldsymbol{B\lambda}) &= 0 \end{aligned}$$

We have

$$\begin{aligned} \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} [\mathbf{x} - \boldsymbol{B\lambda}] = \mathbf{0} \quad &\Leftrightarrow \quad \boldsymbol{B}^\top(\mathbf{x} - \boldsymbol{B\lambda}) = \mathbf{0} \\ &\Leftrightarrow \quad \boldsymbol{B}^\top\boldsymbol{B\lambda} = \boldsymbol{B}^\top\mathbf{x} \end{aligned}$$

Note: $\boldsymbol{B}^\top\boldsymbol{B}$ is invertible $\Rightarrow \boldsymbol{\lambda} = (\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top\mathbf{x}$.

- $\pi_U(\mathbf{x}) = \boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top\mathbf{x} \Rightarrow$ Projection matrix $\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top$.

Example

Example

For a subspace $U = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\} \subseteq \mathbb{R}^3$ and $\mathbf{x} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^3$.

Find

- the coordinates λ of $\mathbf{x}$ in terms of U
- the projection point $\pi_U(\mathbf{x})$
- the projection matrix $\boldsymbol{P}_\pi$.

p. 87; on the black board.

What if $B = (\mathbf{b}_1, \ldots, \mathbf{b}_m)$ is orthonormal?

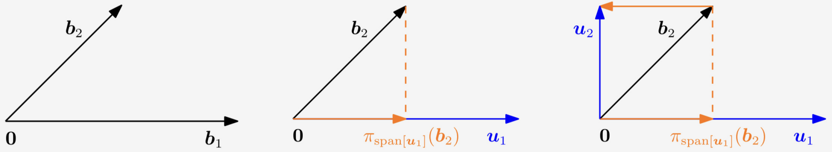
- $\pi_U(\mathbf{x}) = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \mathbf{x} \quad \Rightarrow \pi_U(\mathbf{x}) = \boldsymbol{B}\boldsymbol{B}^\top \mathbf{x}$.
    - $\because \boldsymbol{B}^\top \mathbf{B} = \boldsymbol{I}$.
- Coordinates: $\boldsymbol{\lambda} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \mathbf{x} = \boldsymbol{B}^\top \mathbf{x}$.

Outline

Illustration of Gram-Schmidt Orthogonalization

- **Goal:** Transform any basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of an n-dimensional vector space V into an orthogonal/orthonormal basis of V.

$$\begin{aligned} \mathbf{u}_1 &:= \mathbf{b}_1 \\ \mathbf{u}_k &:= \mathbf{b}_k - \pi_{\text{span}(\{\mathbf{u}_1,\ldots,\mathbf{u}_{k-1}\})}(\mathbf{b}_k), \quad k = 2, \ldots, n. \end{aligned}$$



(a) Original non-orthogonal basis vectors $\boldsymbol{b}_1, \boldsymbol{b}_2$.

(b) First new basis vector $\boldsymbol{u}_1 = \boldsymbol{b}_1$ and projection of $\boldsymbol{b}_2$ onto the subspace spanned by $\boldsymbol{u}_1$.

(c) Orthogonal basis vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2 = \boldsymbol{b}_2 - \pi_{\text{span}[\boldsymbol{u}_1]}(\boldsymbol{b}_2)$.

Example

Example

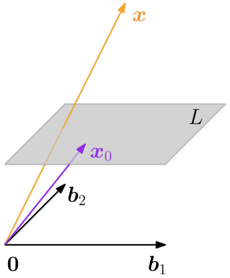
Consider a basis $(\mathbf{b}_1, \mathbf{b}_2)$ of $\mathbb{R}^2$, where $\mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Apply the Gram-Schmidt method to construct an orthonormal basis $(\mathbf{u}_1, \mathbf{u}_2)$ of $\mathbb{R}^2$ (assuming the dot product as the inner product).

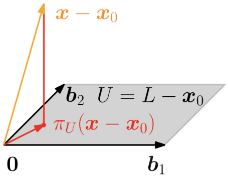
$$
\begin{aligned}
\mathbf{u}_1 &:= \mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \\
\mathbf{u}_2 &:= \mathbf{b}_2 - \pi_{\mathrm{span}(\mathbf{u}_1)}(\mathbf{b}_2) = \mathbf{b}_2 - \frac{\mathbf{u}_1 \mathbf{u_1}^\top}{\|\mathbf{u}_1\|^2} \mathbf{b}_2 \\
&= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{aligned}
$$

Projection onto Affine Spaces

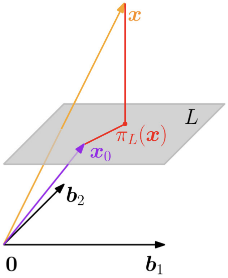
- Given an affine space $L = \mathbf{x}_0 + U$.
  - U is a low-dimensional subspace of V.
- $\pi_L(\mathbf{x}) = \mathbf{x}_0 + \pi_U(\mathbf{x} - \mathbf{x}_0)$



(a) Setting.



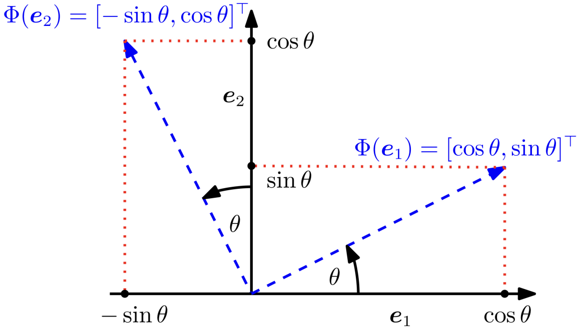
(b) Reduce problem to projection π_U onto vector subspace.



(c) Add support point back in to get affine projection π_L.

Outline

Rotataions in $\mathbb{R}^2$ as An Example



- Standard basis $\mathbf{e} = \{\mathbf{e}_1 = [1 \quad 0]^\top, \quad \mathbf{e}_2 = [0 \quad 1]^\top\}$.
- $\boldsymbol{R}(\theta) = [\Phi(\mathbf{e}_1) \quad \Phi(\mathbf{e}_2)] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Discussions