*** Applied Machine Learning Fundamentals *** Mathematical Foundations

Daniel Wehner

SAP SE

November 11, 2019

Find all slides on GitHub

Lecture Overview

Unit I	Machine Learning Introduction
Unit II	**Mathematical Foundations**
Unit III	Bayesian Decision Theory
Unit IV	Probability Density Estimation
Unit V	Regression
Unit VI	Classification I
Unit VII	Evaluation
Unit VIII	Classification II
Unit IX	Clustering
Unit X	Dimensionality Reduction

Agenda November 11, 2019

Section:

Introduction

Duale Hochschule
Baden-Württemberg

Introduction

Section:

Linear Algebra

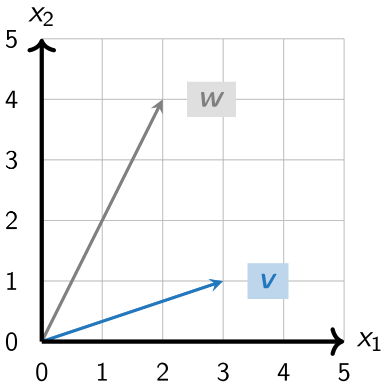
DHBW
Duale Hochschule
Baden-Württemberg

What is a Vector?

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

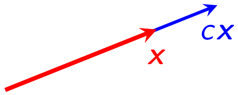
$$\boldsymbol{w} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$



Multiplication by a Scalar

$$c\boldsymbol{x} = c\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix}$$

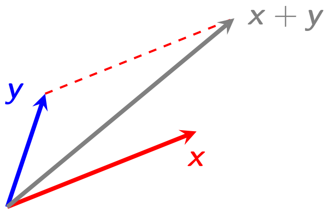
$$2\boldsymbol{v} = 2\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$$



Addition of Vectors

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$



Linear Combination of Vectors

$$\boldsymbol{u} = c_1 \boldsymbol{v}^{(1)} + c_2 \boldsymbol{v}^{(2)} + \cdots + c_n \boldsymbol{v}^{(n)}$$

Vector Transpose and inner and outer Product

- Vector transpose:

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \boldsymbol{v}^\intercal = \begin{bmatrix} 3 & 1 \end{bmatrix}$$

- Inner product / dot product / scalar product:

$$\boldsymbol{v} \cdot \boldsymbol{w} \equiv \boldsymbol{v}^\intercal \boldsymbol{w} \equiv \langle \boldsymbol{v}, \boldsymbol{w} \rangle$$
$$= \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = (3 \cdot 2) + (1 \cdot 4) = 10$$

Vector Transpose and inner and outer Product (Ctd.)

- Outer product:

$$\boldsymbol{v}\boldsymbol{w}^{\mathsf{T}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 12 \\ 2 & 4 \end{bmatrix}$$

The inner product yields a scalar value, the results of an outer product is a matrix!

Length of a Vector

- Length of a vector (Frobenius norm):

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^\intercal \mathbf{x}} \tag{1}$$

$$\|c\mathbf{x}\| = |c| \cdot \|\mathbf{x}\| \tag{2}$$

$$\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\| \tag{3}$$

- Example:

$$\|\mathbf{v}\| = \sqrt{3^2 + 1^2} = 10$$

Angle between Vectors

- The angle between two vectors is given by:

$$\cos \measuredangle(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|} = \frac{\sum_{j=1}^{m} x_j \cdot y_j}{\sqrt{\sum_{j=1}^{m}(x_j)^2} \cdot \sqrt{\sum_{j=1}^{m}(y_j)^2}} \tag{4}$$

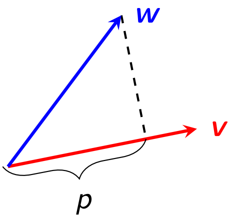
$$\cos \measuredangle(\boldsymbol{v}, \boldsymbol{w}) = \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\|} = \frac{10}{\sqrt{10} \cdot \sqrt{20}} \approx 0.71$$

- Inner product: $\boldsymbol{x} \cdot \boldsymbol{y} = \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\| \cdot \cos \measuredangle(\boldsymbol{x}, \boldsymbol{y})$

Projection of Vectors

- How is the projection of $\boldsymbol{x}$ onto $\boldsymbol{y}$ defined?
- Formally, we have:

$$\begin{aligned} p &= \|\boldsymbol{v}\| \cos \measuredangle(\boldsymbol{v}, \boldsymbol{w}) \\ &= \|\boldsymbol{v}\| \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\|} \\ &= \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{w}\|} \end{aligned} \tag{5}$$



- Note that p is **not a vector!**

What is a Matrix?

General case ($\mathbb{R}^{n \times m}$):

$$\boldsymbol{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix}$$

$$\boldsymbol{M} = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} \qquad \mathbb{R}^{2 \times 3}$$

$$\boldsymbol{N} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbb{R}^{3 \times 3}$$

$$\boldsymbol{P} = \begin{bmatrix} 10 & 1 \\ 11 & 2 \end{bmatrix} \qquad \mathbb{R}^{2 \times 2}$$

Matrix Transpose and Addition

- Transpose of a matrix:

$$\boldsymbol{M}^\intercal = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix}^\intercal = \begin{bmatrix} 3 & 1 \\ 4 & 0 \\ 5 & 1 \end{bmatrix} \tag{6}$$

- Addition of matrices:

$$\boldsymbol{X} + \boldsymbol{Y} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} + \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} X_{11} + Y_{11} & X_{12} + Y_{12} \\ X_{21} + Y_{21} & X_{22} + Y_{22} \end{bmatrix} \tag{7}$$

Matrix Multiplication

- Multiplication by scalars:

$$c\boldsymbol{X} = c\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix} = \begin{bmatrix} c \cdot X_{11} & c \cdot X_{12} & c \cdot X_{13} \\ c \cdot X_{21} & c \cdot X_{22} & c \cdot X_{23} \end{bmatrix} \tag{8}$$

- Matrix-vector multiplication:

$$\boldsymbol{z} = \boldsymbol{X}\boldsymbol{y} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_{11} \cdot y_1 + X_{12} \cdot y_2 \\ X_{21} \cdot y_1 + X_{22} \cdot y_2 \end{bmatrix} \tag{9}$$

Matrix Multiplication (Ctd.)

- Matrix-matrix multiplication:

$$\begin{aligned} \boldsymbol{Z} &= \boldsymbol{X}\boldsymbol{Y} \\ &= \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \\ Y_{31} & Y_{32} \end{bmatrix} \\ &= \begin{bmatrix} X_{11}Y_{11} + X_{12}Y_{21} + X_{13}Y_{31} & X_{11}Y_{12} + X_{12}Y_{22} + X_{13}Y_{32} \\ X_{21}Y_{11} + X_{22}Y_{21} + X_{23}Y_{31} & X_{21}Y_{12} + X_{22}Y_{22} + X_{23}Y_{32} \end{bmatrix} \end{aligned} \tag{10}$$

Matrix Inversion

- Matrix inversion is defined for **square matrices** $\boldsymbol{X} \in \mathbb{R}^{n \times n}$
- A matrix $\boldsymbol{X}$ multiplied by its inverse $\boldsymbol{X}^{-1}$ gives the **identity matrix**:

$$\boldsymbol{X}^{-1}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{I} \tag{11}$$

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{12}$$

- If $\boldsymbol{X}^{-1}$ exists, we say that $\boldsymbol{X}$ is **non-singular**

Matrix Inversion (Ctd.)

- It holds that ($\boldsymbol{C}$ is the cofactor matrix):

$$\boldsymbol{X}^{-1} = \frac{1}{\det(\boldsymbol{X})} \boldsymbol{C}^{\intercal} \tag{13}$$

- A condition for invertability is that **the determinant has to be different than zero**
- **Example:**

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \det(\boldsymbol{X}) = 0 \qquad \boldsymbol{X}^{-1} = ?$$

Matrix Inversion Example

$$X = \begin{bmatrix} 1 & 1/2 \\ -1 & 1 \end{bmatrix} \qquad X^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ 2/3 & 2/3 \end{bmatrix}$$

Please verify!

$$XX^{-1} = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = X^{-1}X$$

Use for example the Gauss-Jordan algorithm to find the inverse!

Matrix Pseudoinverse

- **Question:** How can we invert a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ which is not squared?
- **Left pseudoinverse $\boldsymbol{X}^{\#}\boldsymbol{X}$:**

$$\boldsymbol{X}^{\#}\boldsymbol{X} = \underbrace{(\boldsymbol{X}^{\intercal}\boldsymbol{X})^{-1}\boldsymbol{X}^{\intercal}}_{\text{left-multiplied}}\boldsymbol{X} = \boldsymbol{I}_m \tag{14}$$

- **Right pseudoinverse $\boldsymbol{X}\boldsymbol{X}^{\#}$:**

$$\boldsymbol{X}\boldsymbol{X}^{\#} = \boldsymbol{X}\underbrace{\boldsymbol{X}^{\intercal}(\boldsymbol{X}\boldsymbol{X}^{\intercal})^{-1}}_{\text{right-multiplied}} = \boldsymbol{I}_n \tag{15}$$

Eigenvectors and Eigenvalues

- Some vectors $\boldsymbol{v}$ only change their length when multiplied by a matrix $\boldsymbol{X}$

Symmetric Matrices

- A squared $n \times n$ matrix $\boldsymbol{X}$ is **symmetric**, iff

$$\forall i, j: \qquad X_{ij} = X_{ji} \tag{16}$$

$$\boldsymbol{X} = \boldsymbol{X}^{\intercal} \tag{17}$$

- Some properties:
  - The inverse $\boldsymbol{X}^{-1}$ is also symmetric
  - **Eigen-decomposition:** $\boldsymbol{X}$ can be decomposed into $\boldsymbol{X} = \boldsymbol{QDQ}^{\intercal}$, where the columns of $\boldsymbol{Q}$ are the eigenvectors of $\boldsymbol{X}$, and $\boldsymbol{D}$ is a diagonal matrix whose entries are the corresponding eigenvalues

Positive (semi-)definite Matrices

- A **squared symmetric** matrix $\boldsymbol{X}^{n \times n}$ is positive definite, iff for any vector $\boldsymbol{y} \in \mathbb{R}^n$:

$$\boldsymbol{y}^\intercal \boldsymbol{X} \boldsymbol{y} > 0 \tag{18}$$

- Or positive semi-definite, iff $\boldsymbol{y}^\intercal \boldsymbol{X} \boldsymbol{y} \geqslant 0$

Such matrices are important in machine learning. For instance, the covariance matrix is always positive semi-definite.

Section:

Statistics

Duale Hochschule
Baden-Württemberg

Section:

Optimization

Section:

Wrap-Up

Summary

-

Important

Self-Test Questions

1.

What's next...?

Unit I	Machine Learning Introduction
Unit II	Mathematical Foundations
Unit III	**Bayesian Decision Theory**
Unit IV	Probability Density Estimation
Unit V	Regression
Unit VI	Classification I
Unit VII	Evaluation
Unit VIII	Classification II
Unit IX	Clustering
Unit X	Dimensionality Reduction

Recommended Literature and further Reading

Thank you very much for the attention!

Topic: *** Applied Machine Learning Fundamentals *** Mathematical Foundations
Date: November 11, 2019

Contact:
Daniel Wehner
SAP SE
daniel.wehner@sap.com

Do you have any questions?