Linear Algebra for Machine Learning in Python

Dr. Moritz Wolter

August 10, 2022

High Performance Computing and Analytics Lab

Overview

Introcution

Essential operations

Linear curve fitting

Regularization using eigen- oder singular values

Introcution

Motvating linear algebra

TODO

Matrices

$\mathbf{A} \in \mathbb{R}^{m,n}$ is a real-valued Matrix with m rows and n columns.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, a_{ij} \in \mathbb{R}. \tag{1}$$

Essential operations

Addition

To matrices $\mathbf{A} \in \mathbf{R}^{m,n}$ and $\mathbf{B} \in \mathbf{R}^{m,n}$ can be added by adding their elements.

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{pmatrix} \quad (2)$$

Multiplication

Multiply $\mathbf{A} \in \mathbb{R}^{m,n}$ by $\mathbf{B} \in \mathbb{R}^{n,p}$ produces $\mathbf{C} \in \mathbb{R}^{m,p}$,

$$\mathbf{AB} = \mathbf{C}. \tag{3}$$

To compute $\mathbf{C}$ the elements in the rows of $\mathbf{A}$ are multiplied with the column elements of $\mathbf{C}$ and the products added,

$$c_{ik} = \sum_{j=1}^{m} a_{ij} \cdot b_{jk}. \tag{4}$$

Multiplication

Multiply $\mathbf{A} \in \mathbb{R}^{m,n}$ by $\mathbf{B} \in \mathbb{R}^{n,p}$ produces $\mathbf{C} \in \mathbb{R}^{m,p}$,

$$\mathbf{AB} = \mathbf{C}. \tag{3}$$

To compute $\mathbf{C}$ the elements in the rows of $\mathbf{A}$ are multiplied with the column elements of $\mathbf{C}$ and the products added,

$$c_{ik} = \sum_{j=1}^{m} a_{ij} \cdot b_{jk}. \tag{4}$$

Define on the board:

- Dot product $\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$ for two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.
- Row times column view [Str+09]:

Matrix inverse

TODO

The Transpose

TODO: use an orthogonal matrix!

Motivation of the determinant

TODO

Computing determinants in two or three dimensions

The two dimensional case:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21} \tag{5}$$

$$\tag{6}$$

Computing the determinant of a three dimensional matrix.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \cdot \begin{vmatrix} a_{21} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \cdot \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \cdot \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \tag{7}$$

Computing determinants in two or three dimensions

The two dimensional case:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21} \tag{5}$$

$$\tag{6}$$

Computing the determinant of a three dimensional matrix.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \cdot \begin{vmatrix} a_{21} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \cdot \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \cdot \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \tag{7}$$

Draw the sign pattern on the board:

$$\begin{vmatrix} + & - & + & \dots \\ - & + & - & \dots \\ + & - & + & \dots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} \tag{8}$$

The determinant can be expandend along any column as long as the sign pattern is respected.

Determinants in n-dimensions

$$\begin{vmatrix} a_{11} & a_{21} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{m2} & \dots & a_{mn} \end{vmatrix} + a_{21} \begin{vmatrix} a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{m2} & \dots & a_{mn} \end{vmatrix}$$

$$-a_{m1} \begin{vmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \end{vmatrix}$$

Linear curve fitting

What is the best line connecting measurements?

The Pseudoinverse

$$\mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \tag{9}$$

The Pseudoinverse

$$\mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \tag{9}$$

Sometimes solving $\mathbf{Ax} + \mathbf{b} = 0$ is implossible. One the board, derive:

$$\min_x \frac{1}{2}|\mathbf{Ax} - \mathbf{b}|^2 \tag{10}$$

$$\tag{11}$$

At the optimum we expect,

$$\begin{aligned}
0 &= \nabla_x \frac{1}{2}|\mathbf{Ax} - \mathbf{b}|^2 && (12)\\
&= \nabla_x \frac{1}{2}(\mathbf{Ax} - \mathbf{b})^T(\mathbf{Ax} - \mathbf{b}) && (13)\\
&= (\mathbf{Ax} - \mathbf{b})\mathbf{A}^T && (14)\\
&= \mathbf{A}^T\mathbf{Ax} - \mathbf{A}^T\mathbf{b} && (15)\\
\mathbf{A}^T\mathbf{b} &= \mathbf{A}^T\mathbf{Ax} && (16)\\
(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} &= \mathbf{x} && (17)
\end{aligned}$$

Regularization using eigen- oder singular values

Eigenvalue-Decomposition

TODO

Singular-Value-Decomposition

TODO

References

[Str+09] Gilbert Strang, Gilbert Strang, Gilbert Strang, and Gilbert Strang. *Introduction to linear algebra*. Vol. 4. Wellesley-Cambridge Press Wellesley, MA, 2009.