Linear Algebra Review and NumPy Basics[1]

Chris J. Maddison

University of Toronto

[1] Slides adapted from Ian Goodfellow's *Deep Learning* textbook lectures

About this tutorial

- Not a comprehensive survey of all of linear algebra.
- Focused on the subset most relevant to machine learning.
- Larger subset: e.g., Linear Algebra by Gilbert Strang

Scalars

- A scalar is a single number
- Integers, real numbers, rational numbers, etc.
- Typically denoted in italic font:

$$a, n, x$$

Vectors

- A vector is an array of d numbers
- x_i be integer, real, binary, etc.
- Notation to denote type and size:

$$\boldsymbol{x} \in \mathbb{R}^d$$

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

Matrices

- A matrix is an array of numbers with two indices
- $A_{i,j}$ be integer, real, binary, etc.
- Notation to denote type and size:

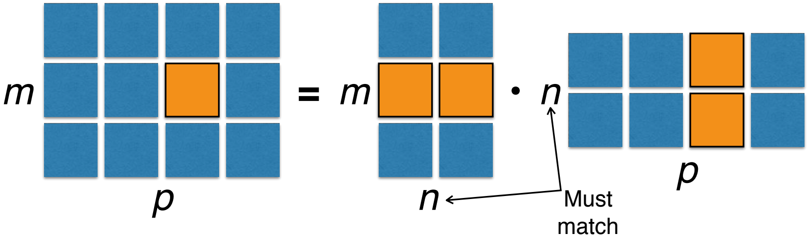
$$A \in \mathbb{R}^{m \times n}$$

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

Matrix (Dot) Product

Matrix product AB is the matrix such that

$$(AB)_{i,j} = \sum_k A_{i,k} B_{k,j}.$$



(Goodfellow 2016)

This also defines matrix-vector products $A\boldsymbol{x}$ and $\boldsymbol{x}^\top A$.

Identity Matrix

The identity matrix for $\mathbb{R}^d$ is the matrix I_d such that

$$\forall \boldsymbol{x} \in \mathbb{R}^d,\ \ I_d \boldsymbol{x} = \boldsymbol{x}$$

For example, I_3:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matrix Transpose

The transpose of a matrix A is the matrix $A^\top$ such that $(A^\top)_{i,j} = A_{j,i}$.

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \implies A^\top = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

The transpose of a matrix can be thought of as a mirror image across the main diagonal. The transpose switches the order of the matrix product.

$$(AB)^\top = B^\top A^\top$$

Systems of equations

The matrix equation

$$A\boldsymbol{x} = \boldsymbol{b}$$

expands to

$$
\begin{aligned}
A_{1,1}x_1 + A_{1,2}x_2 + \cdots A_{1,n}x_n &= b_1 \\
A_{2,1}x_1 + A_{2,2}x_2 + \cdots A_{2,n}x_n &= b_2 \\
&\vdots \\
A_{m,1}x_1 + A_{m,2}x_2 + \cdots A_{m,n}x_n &= b_m
\end{aligned}
$$

Solving Systems of Equations

A linear system of equations can have:

- No solution
- Many solutions
- Exactly one solution, i.e. multiplying by the matrix is an invertible function

Matrix Inversion

The matrix inverse of A is the matrix A^{-1} such that

$$A^{-1}A = I_d$$

Solving a linear system using an inverse:

$$\begin{aligned} A\boldsymbol{x} &= \boldsymbol{b} \\ A^{-1}A\boldsymbol{x} &= \boldsymbol{b} \\ I_d\boldsymbol{x} &= A^{-1}\boldsymbol{b} \end{aligned}$$

Can be numerically unstable to implement it this way in the computer, but useful for abstract analysis.

Invertibility

Be careful, the matrix inverse does not always exist. For example, a matrix cannot be inverted if...

- More rows than columns
- More columns than rows
- Rows or columns can be written as linear combinations of other rows or columns ("linearly dependent")

Norms

- A norm is a function that measures how "large" a vector is
- Similar to a distance between zero and the point represented by the vector

$$
\begin{aligned}
&f(\boldsymbol{x}) = 0 \implies \boldsymbol{x} = 0 \\
&f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y}) \text{ (the triangle inequality)} \\
&\forall a \in \mathbb{R},\ f(a\boldsymbol{x}) = |a| f(\boldsymbol{x})
\end{aligned}
$$

Norms

- L^p norm

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

- Most popular norm: L2 norm, $p = 2$, i.e., the Euclidean norm.
- L1 norm:

$$\|\boldsymbol{x}\|_1 = \sum_i |x_i|$$

- Max norm, infinite norm:

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i|$$

Special Matrices and Vectors

- Unit vector:

$$\|\boldsymbol{x}\|_2 = 1$$

- Symmetric matrix:

$$A = A^\top$$

- Orthogonal matrix

$$A^\top A = AA^\top = I_d$$
$$A^\top = A^{-1}$$

Trace

- The trace of an $n \times n$ matrix is the sum of the diagonal

$$\mathrm{Tr}(A) = \sum_i A_{i,i}$$

- It satisfies some nice commutative properties

$$\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB) = \mathrm{Tr}(BCA)$$

How to learn linear algebra

- Lots of practice problems.
- Start writing out things explicitly with summations and individual indexes.
- Eventually you will be able to mostly use matrix and vector product notation quickly and easily.

NumPy

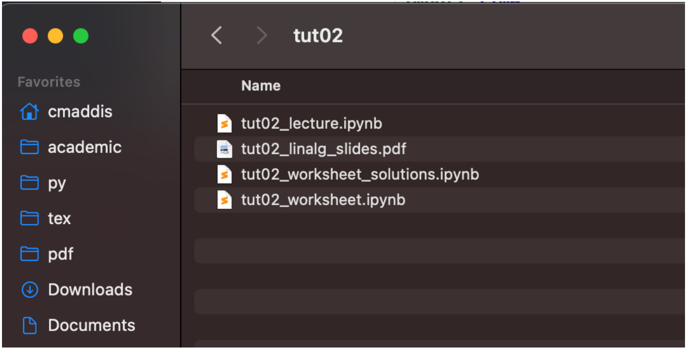
- NumPy is a software package written for the Python programming language the helps us perform vector-matrix operations very efficiently.
- We will be running through some examples today to get a sense of how to use NumPy.
- First, we will show you how to open a Python Jupyter Notebook in the UofT Jupyter Hub.

Note

This tutorial is completely optional, we do not expect you to be able to use Python Jupyter Notebooks for any assessments.

Python Jupyter Notebooks (.ipynb)

.ipynb files are python scripts that organize code in to runnable cells (will see examples today)



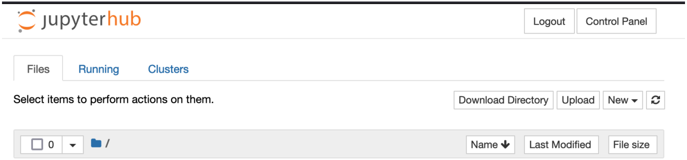
Can run them on UofT Jupyter Hub or Google Colab.

Python Jupyter Notebooks

UofT Jupyter Hub: go to `https://jupyter.utoronto.ca` and log in.

Python Jupyter Notebooks

When you log in, you should see this:

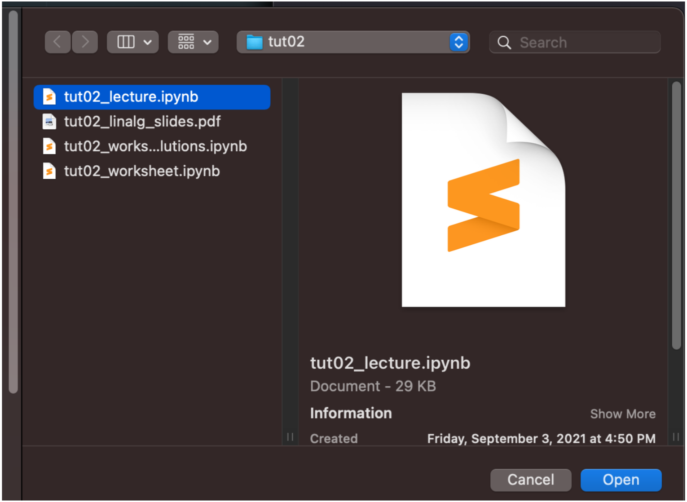


Python Jupyter Notebooks

To load a Jupyter Notebook, click on the Upload button:

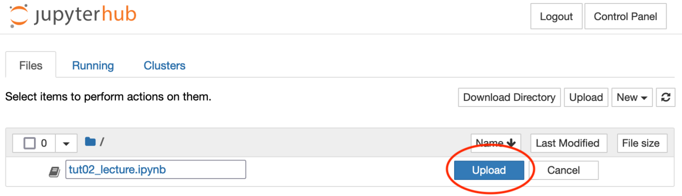
Python Jupyter Notebooks

You will get a dialog. Select the .ipynb you want and open it:



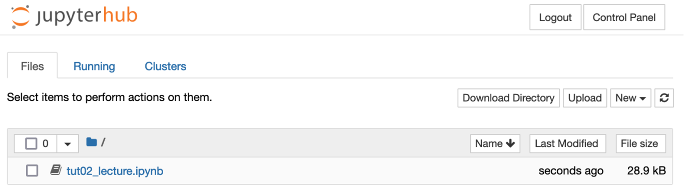
Python Jupyter Notebooks

Finally, click the blue Upload button:



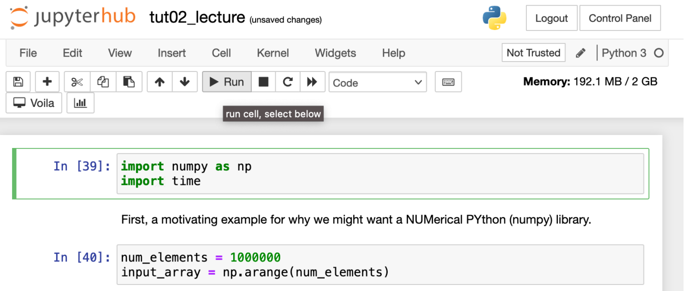
Python Jupyter Notebooks

Now you should have it uploaded!



Python Jupyter Notebooks

Click on the link to open the Notebook.



Now we will show you how to run things.