Tutorial 1

Statistical Computation and Analysis

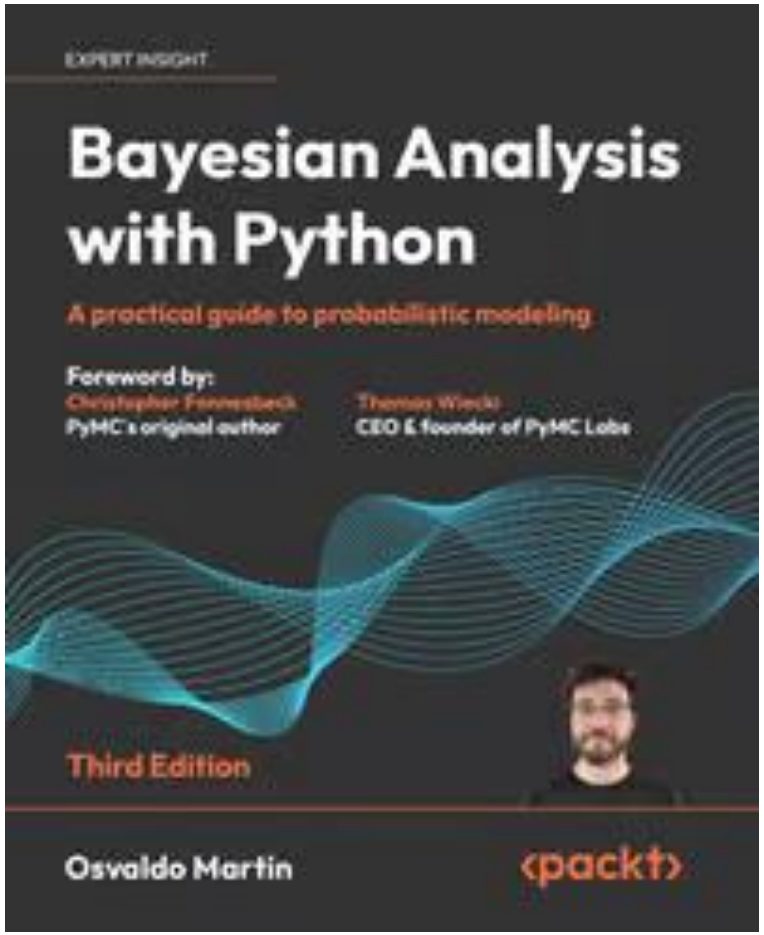
Spring 2025

Course Information

- The course is in English.
- The homework assignments need to be handed in in English.
- The course is in Python.
  - The codes for the examples in the lectures and tutorials will be made available on Moodle.
- Course grade:
  - Pre-lecture questions (5%)
  - Homework assignments (15%)
    - 5 assignments (3.6% per exercise)
  - Midterm quiz (10%, magen)
    - 6.5.25
    - Open moodle except previous tests and quizzes
    - 1 hour
  - Exam (70%)
    - On computers on campus
    - Open moodle except previous tests and quizzes
    - 2 hours

Bayesian Analysis with Python

- 3rd edition
- Publisher's website
- Full text available
  - the library



Tutorial Outline

- Python
  - Libraries
  - Google colab

Practice Google Colab Notebook

- **Python basics**

https://colab.research.google.com/github/cs231n/cs231n.github.io/blob/master/python-colab.ipynb

- **Code for this tutorial:**

Moodle – Tutorial1_2025.ipynb

Data Types

- **Integers**
- **Floating-point numbers**
- **Boolean**

MATLAB	PYTHON
true	True
false	False
&	and
|	or
~	not
~=	!=

Data Types

- **Strings**
  - Concatenation: ‘hello’ + str(var) + ‘_’ + ‘world’
  - fstring: f‘hello{var }_world’
- **Indexing:**

MATLAB	PYTHON
Starts from 1	Starts from 0
array(1:3)	array[0:2]

Containers

- **Lists:** ordered sequences of elements. They can contain any combination of data types and are defined using square brackets, e.g., l = [1, 2, "three"].
- **Tuples:** similar to lists, but they are immutable (cannot be changed). They are defined using parentheses, e.g., t = (1, 2, "three").
- **Dictionaries:** used to store data values in key:value pairs. They are defined using curly brackets, e.g., d = {"brand": "Ford", "model":"Mustang“}.

For Loops

- for i in range(5):

```
    print(i)
```

Output:

0

1

2

3

4

For Loops

- animals = ['cat', 'dog', 'monkey’] *(list)*

```
for animal in animals:
    print(animal)
```

Output:

cat

dog

monkey

For Loops

- for i, animal in enumerate(animals):

```
    print(f"Number {i}, animal {animal}")
```

Output:

Number 0, animal cat

Number 1, animal dog

Number 2, animal monkey

Functions

- def function_name(input arguments):
- Defined before function call / separate code and imported.

Python Libraries

- **NumPy:** provides support for arrays, matrices, and mathematical operations on them.
- **Pandas:** provides tools for data manipulation and analysis, including data structures and functions for transforming, cleaning, and analyzing datasets.
- **SciPy:** builds on NumPy and provides additional functionality for scientific computing, including statistical functions.
- **Matplotlib:** data visualization.
- **Seaborn:** more data visualization.
- **Math:** math operations.

Python Libraries

- **Import libraries and external functions:**
  - import numpy as np (np.array([1, 2, 3])).
  - from file_name import function_name
  - from numpy import * - imports all numpy functions, not recommended.

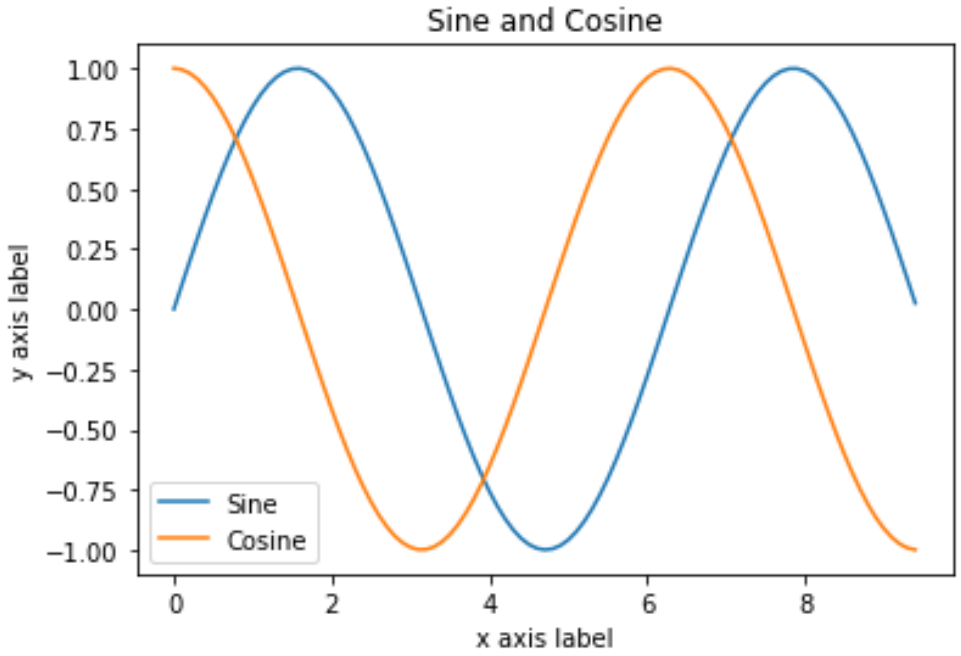
Matplotlib

- **Plots:**

```
import matplotlib.pyplot as plt
import numpy as np

x = np.arange(0, 3 * np.pi, 0.1)
y_sin = np.sin(x)
y_cos = np.cos(x)

# Plot the points using matplotlib
plt.plot(x, y_sin)
plt.plot(x, y_cos)
plt.xlabel('x axis label')
plt.ylabel('y axis label')
plt.title('Sine and Cosine')
plt.legend(['Sine', 'Cosine'])
```



Pandas

- Provides two types of classes for handling data:
  - Series – a 1D labelled array holding data of any type.
  - DataFrame – a 2D data structure that holds data like a two dimension array or a table with rows and columns.

Xarray

- Enables working with multi-dimensional dataframes.
  - Coordinates: the coordinates at which the measurements were taken.
    - For example, x value, y value and times.
    - Lat and lon come from the words latitude and longitude, and are like the rows and the columns in a dataframe.
    - Here we have an additional dimension of time, for example.
    - There can be more dimensions.
  - Data variables: what was measured at the coordinates.
    - There can be one, or more than one.
  - Indexes: refer to the coordinate values used for efficient data selection.
    - In most cases they are created from the coordinates, but there are cases in which they differ.
  - Attributes: extra metadata about the dataset of the data variables.
    - Units
    - Sources

Xarray

- Extracting data
  - Label based
  - Position based

Homework Template

- HomeworkTemplate.ipynb on the moodle