PHY 835: Exercise 1
Released Jan 26, 2021;Target Due Date: Feb 9, 2021

1. Getting started

The aim of this exercise is to get yourself set up with an environment that allows you to run certain packages which are essential to perform future exercises. Please ensure that you have the following packages installed:

1. python 3
2. Packages: numpy, matplotlib, scikit-learn, jupyter notebooks
3. Tensorflow

Anaconda.org is a user friendly platform for installing these software packages.

To check your results, please execute the Juypter notebook getting_started.ipynb.

2. Visualizing data

Familiarize yourself with matplotlib, an excellent library to visualize data. Ability to visualize your results efficiently is key in data science and data analysis.

1. Plot $f(x) = x^2$ for the interval $(-5, 5)$.
2. Generate a dataset of 100 points drawn at random from a 2D Gaussian with mean $(1, 0.5)$ and standard deviation $(0.4, 0.4)$.
3. Now, highlight the mean and 1-σ contour lines in a plot with your data points.

Note that matplotlib among other python based packages which we will use in this course have been vital to produce the images released from the EHT collaboration.

3. Linear regression

Here we familiarize ourselves with linear regression of simple polynomial models in python.

1. Take the function $f(x) = 7.2 - 3.1x + 4.3x^3$. For 100 equally spaced bins in the interval $(-5, 8)$ add a Gaussian noise (independent in each bin) with respective means 0 and standard deviation $\sigma = 150$ to the data points.
2. Now fit polynomials of degree 1,2,3,5,10 to the entire dataset, using linear regression with ordinary least squares.
3. Repeat the same fitting but with reduced data sets. Visualize your findings and describe the differences? How do they depend on the interval you pick the data from?
4. (Optional and more challenging): Take a cubic polynomial but now in terms of a complex variable and complex coefficients. Repeat steps a) and b) with the modified function. Output your respective best-fit polynomial.