

Exercise Sheet 2

Exercise 4

Load the data provided in `data_KDE_1D.csv` and:

- (a) Explore the data using suitable methods and tools.
- (b) Group the data to generate a meaningful histogram of the data.
- (c) Try to fit a normal distribution into the data. Plot the fitted normal distribution on top of the histogram.
Hint: use `scipy.stats.norm.fit` to estimate the parameters of the normal distribution.
- (d) Multivariate kernel density estimation: Put a kernel (Gaussian, box, triangular) on top of every data point. Discuss the results.
Hint: you can use python libraries such as `KDEpy` for your solution.
- (e) Vary the parameters of the kernels (bandwidth, weights, ...) used in (d) and combine them to get a KDE.

Exercise 5

Load the data provided in `data_KDE_2D.csv` and:

- (a) Explore the data using suitable methods and tools.
- (b) Group the data to generate a meaningful histogram of the data.
- (c) Try to fit a normal distribution into the data. Plot the surface and contours of the fitted normal distribution on top of the data points.
Hint: Calculate the mean vector and the covariance matrix of the data. To calculate the PDF of the multivariate ND, use `scipy.stats.multivariate.normal`.
- (d) Multivariate kernel density estimation: Put a kernel (Gaussian, box, triangular) on top of every data point. Discuss the results.
Hint: you can use python libraries such as `KDEpy` for your solution.
- (e) Vary the parameters of the kernels (bandwidth, weights, ...) used in (d) and combine them to get a KDE.

Exercise 6

Load the data provided in `data_KDE_GM.csv` and:

- (a) Use `scipy.stats.gaussian_kde` to perform a KDE on the data.
- (b) Use `sklearn.mixture.GaussianMixture` to fit a gaussian mixture model for the provided data. Vary the number of components starting with 1 up to 5.
- (c) Visualize and discuss your results