

Assignment #1

Submission type: individual

Deadline: October 4th, 2020 (end-of-the-day)

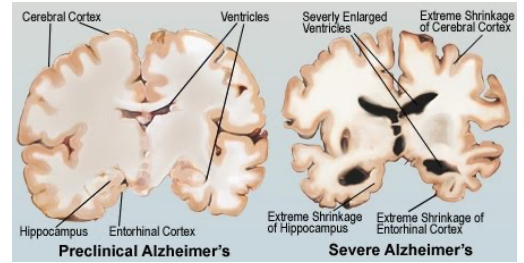
Purpose of assignment:

- Getting familiar with a Colab environment and use of python libraries
- Understanding univariate/multivariate Gaussian distribution
- Learning data analysis

Write a code to compare and analyze features by diagnostic group using a univariate/bivariate Gaussian estimation.

Read All Instructions Carefully

1. Dataset in “alzheimers_disease_dataset.csv”
 - Column A: Subject ID
 - Column B: Diagnosis group
0: Cognitively Normal (CN)
1: Mild Cognitive Impairment (MCI)
2: Alzheimer’s Disease (AD)
 - Column C: cortical volume of a hippocampus
 - Column D: cortical volume of an entorhinal cortex



2. Maximum Likelihood Estimation (MLE) is a method that determines values for the parameters of the model. The probability density of observing a single data point, that is generated from a Gaussian distribution is given by:

Univariate MLE

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (1)$$
$$\hat{\mu} = \frac{1}{m} \sum_i^m x_i, \quad \hat{\sigma}^2 = \frac{1}{m} \sum_i^m (x_i - \hat{\mu})^2$$

Multivariate MLE

$$P(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})^T\right) \quad (2)$$
$$\hat{\boldsymbol{\mu}} = \frac{1}{m} \sum_i^m \mathbf{x}_i, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{m} \sum_i^m (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T$$

where m is the number of samples/observations.

3. **Task Description:** Compute and plot the univariate and bivariate Gaussian distribution for each symptom-based diagnosis group.
 - In other words, compute and plot Gaussian distributions:
 - i. Feature 1 (Hippocampus) univariate Gaussian distribution per group
 - ii. Feature 2 (Entorhinal cortex) univariate Gaussian distribution per group
 - iii. Hippocampus-Entorhinal cortex bivariate Gaussian distribution per group
 - Example with Scipy:

http://web.vu.lt/mif/a.buteikis/wp-content/uploads/PE_Book/3-4-UnivarMLE.html

4. Implementation instructions

- Use the Google Colab (<https://colab.research.google.com/>)
- **Only allowed libraries are pandas, numpy, matplotlib (or any other visualization libraries such as Pillow).**

[Submission: “BNCS411_ **yourID_name**.ipynb” file (in Colab menu: File-.ipynb Download)]

1. For a segment of codes, you should describe their function by using “markdown” scripts.
2. It is also required to write your analysis in regard to the resulting figures.

TA’s note: Beware of plagiarism (we use proprietary plagiarism checkers in addition to BlackBoard plagiarism checker). Ask the TAs if you have any questions.