

Sanuraag Mishra

Find Me Out

18 May 2023

Week - 2

What I learnt this week :-

- ☑ machine learning includes humongous amounts of data, multiple hyperparameters, and a complex environment, uncertainties are bound to exist
- ☑ frequentist probability vs Bayesian probability
- ☑ Discrete and Continuous Variables
- ☑ probability mass function (PMF) vs probability density function (PDF)
- ☑ marginal probability distribution, conditional probability distribution.

☑ Joint probability expression
$$P(x_1, \dots, x_n) = P(x_1) \prod_{i=2}^n P(x_i | x_1, \dots, x_{i-1})$$

- ☑ events are conditionally independent, the probability is given as

$$P(x = X, y = Y | z = Z) = P(x = X | z = Z) * P(y = Y | z = Z)$$

☑ Bayes Rule
$$P(y|x) = \frac{P(x|y) * P(y)}{P(x)}$$

☑ Expected Value
$$E_{x \sim P}[f(x)] = \sum_x P(x) f(x)$$

☑ conditional expected value
$$E_{x \sim P}[f(x)] = \sum_x P(x|y) f(x)$$

☑ Variance
$$Var(f(x)) = E[(f(x) - E[f(x)])^2]$$

☑ Covariance
$$Cov(f(x_1), f(x_2)) = E[(f(x_1) - E(f(x_1)))(f(x_2) - E(f(x_2)))]$$

- ☑ Various types of Distributions:

Bernoulli, Binomial, Multinoulli, Multinomial, Gaussian, Exponential

- ☑ Intuitive, mathematical understanding of maximum likelihood estimation (MLE).

- ☒ Difference between MLE and MSE(mean squared error).
- ☒ MLE calculation for a Gaussian and binomial distribution.

NUMPY YT

- ☒ Numpy arrays Slicing and Index
- ☒ Reshaping Arrays
- ☒ Stacking & Splitting
- ☒ Copying Arrays
- ☒ Universal Math
- ☒ Reading From Files
- ☒ Statistics Functions
- ☒ Creating Formulas
- ☒ Trigonometry Functions
- ☒ Linear Algebra
- ☒ Saving & Loading NumPy Objects
- ☒ Loading Libraries in Anaconda
- ☒ Financial Functions
- ☒ Comparison Functions

PANDAS YT

- ☐ Series
- ☒ Creating DataFrames
- ☒ Editing / Retrieving Data
- ☒ Conditional Selection
- ☒ Importing Data from Anywhere
- ☒ Basics & Math
- ☒ Grouping Data
- ☒ Concatenate, Merge & Join Data
- ☒ Statistics
- ☒ Iteration
- ☒ Sorting
- ☒ Manipulate Data with Functions
- ☒ Aligning, Reindexing, Renaming
- ☒ MultiIndex
- ☒ Cleaning Data
- ☒ Real World Examples
- ☒ Visualization
- ☒ Installation & Virtual Environments

MATPLOTLIB Yt

- ☒ Importing
- ☒ Simple Plotting

- ☒ Multiple Plots
- ☒ Using Figure Objects
- ☒ Subplots
- ☒ Appearance Options
- ☒ Saving Plots
- ☒ Working with Pandas Dataframe
- ☒ TeX Markup (Math Symbols)
- ☒ Histograms
- ☒ Bar Charts
- ☒ Pie Charts
- ☒ Timeseries
- ☒ Tables
- ☒ Scatterplots
- ☒ 3D Surfaces
- ☒ Matplotlib Finance
- ☒ Heatmaps

What I implemented this week:-

- *Downloaded Iris.csv from Kaggle*
- *Printed the name of column along with data type using for loop*
- Found the mean and variance of each column.
- Plotted histograms for each column by choosing the right number of bins
- correlation matrix/heat map
- Plotted scatterplot between each pair of columns
- Q-Q plots
- Box and Whisker's plot

Link to my Week 2 collab

<https://colab.research.google.com/drive/1OCA6MQSCDFptXL9ew8jO9p80LjbuxuGU?usp=sharing>