

ReadMe - Horse Kick Death Prediction (Q3)

General Instruction:

Dependencies:

- 1) Numpy
- 2) Math
- 3) Matplotlib
- 4) Scipy
- 5) Warnings

Running code:

1. Run in Terminal -

To run the code, run the following command in the terminal

Code: python3 Assignment1_Q3.py

2. **Run in Google Colab - (Highly Recommended)**

Upload the '**Assignment-1_Q4.ipynb**' file in Google colab and run

3. **Run in Local Machine (Jupyter Notebook)**

It will show indentation error as code was done in Google Colab where indentation is of 2 spaces and in Local Machine Jupyter Notebook the indentation is of 4 spaces. So it is highly recommended to run the ipynb file in Google Colab.

Dataset Loading:

Dataset is already loaded in the python script as well as in the Jupyter Notebook as a numpy array and are splitted into train and test sets.

Description of Functions:

1. Estimate_Parameters(dataset,mode):

Function to compute Estimate of Poisson Parameter

Args: dataset = Train Dataset

mode = 'MLE' for Maximum Likelihood Estimate

'MAP' for Maximum a-posteriori Estimate

Output: If mode = 'MLE' then, Parameter of Poisson Distribution

If mode = 'MAP' then, Parameter of Poisson Distribution

2. Predict(estimated_param,years):

Given MLE Parameters of Poisson Distribution and Years of Prediction this function predicts the number of deaths for the next "Years" years

Args: estimated_param = MLE or MAP estimated Parameter of Poisson

Years = Years of Prediction

Output: Predicted Deaths, Shape = (Num_Corps,Years)

3. RMSE(x_true,x_pred):

Root Mean Squared Error

Args: x_true = true test data

x_pred = pred test data

Output: RMSE for every corps

Graph: RMSE value for each corps

4. plot_Modes(X_train,X_test,Corps_list,mode_label):

Function to plot actual modes of the dataset

Args: X_train: Training Dataset

X_test: Actual Testing Dataset

Corps_list = Corps Numbers as a list to plot mode

mode_label = Plot Label (Distribution of Actual Mode or Predicted Mode)

Output: Plots for distribution and mode for the corps

5. Prussian_Visualize(train_dataset):

Function to visualize Prussian Death Train Dataset

Args: Train dataset

Output: Plots for deaths of each corps over years

6. plot_probs(dataset,corps_number):

Function to plot likelihood, prior and posterior of a given corps

Args: dataset = Train Dataset

corps_number = corps number for which plots are required

Output: Likelihood,Prior and Posterior plot for the corps

Modes of Prior, Likelihood and Posterior (Which comes out to be equal to MLE and MAP of Poisson Parameter Lambda)

7. plot_data(data, label, xlab, ylab):

Utility function to plot various data values

Args: data = Values to plot

label = Label of the graph

xlab = X-axis label

ylab = Y-axis label

Output: Graph of data values

8. `plot_loss_compare(loss1,loss2):`

Function to plot loss values and compare

Args: loss1 = Loss Values of prediction using MLE

loss2 = Loss Values of prediction using MAP

Output: Graph of RMSE Loss Value of Predictions using MLE and MAP

9. `MLE_MAP_Visualize(train_dataset,mle,map):`

Function to visualize Prussian Death Train Dataset

Args: Train dataset,

mle = MLE of poisson parameters for all the corps

map = MAP of poisson parameters for all the corps

Output: MLE and MAP values for each corps with respect to corresponding corps data