

# PyTorch

Saurabh Burewar (B18CSE050)

## Dataset

Link to the dataset: <https://www.kaggle.com/c/titanic/data>

The dataset shows different attributes of the passengers of Titanic and whether they survived or not. I have only used “train.csv” and divided this file itself into train and test sets with a 70:30 size ratio. It contains the following features -

| Features                 | Definition   | Key  |
|--------------------------|--|--|
| Pclass                   | Ticket class   | 1=Upper, 2=Middle, 3=Lower                           |
| Sex                      | Sex  | male, female   |
| Age                      | Age in years   |  |
| Sibsp                    | # of siblings/spouses aboard the Titanic               |  |
| Parch                    | # of parents/children aboard the Titanic               |  |
| Ticket                   | Ticket number  |  |
| Fare                     | Passenger fare   |  |
| Cabin                    | Cabin number   |  |
| Embarked                 | Port of embarkation                                    | C = Cherbourg,<br>Q = Queenstown,<br>S = Southampton |
| Survived (Target)        | Survived or not  | 0 = No, 1 = Yes                                      |
| sex_factor (added by me) | Factorize the feature “Sex” to convert it to numerical | 0 = male, 1 = female                                 |
| em_factor (added by me)  | Factorize the feature                                  | 0 = S = Southampton,                                 |

|  |                                       |  |
|--|---------------------------------------|--|
|  | "Embarked" to convert it to numerical | 1 = C = Cherbourg,<br>2 = Q = Queenstown |
|--|---------------------------------------|--|

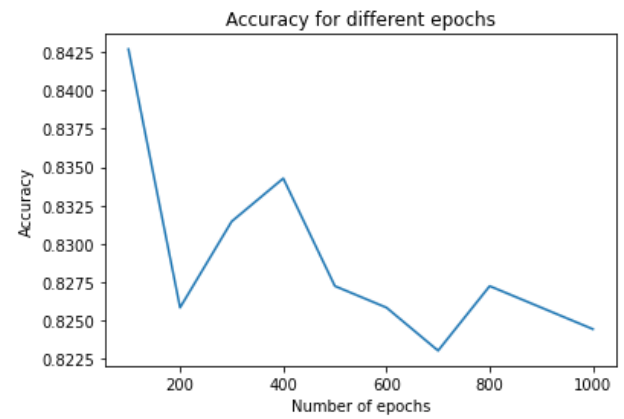
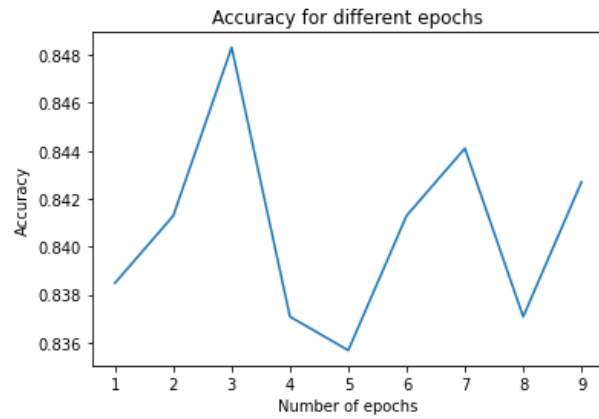
- Features used in the training of the models
- Target variable

## Experiments

The following table shows the accuracy increase with iterations-

| Number of iterations | Accuracy (in %)    | Number of iterations | Accuracy (in %)    |
|----------------------|--------------------|----------------------|--------------------|
| 10                   | 0.8356741573033708 | 100                  | 0.8426966292134831 |
| 20                   | 0.8286516853932584 | 200                  | 0.8258426966292135 |
| 30                   | 0.827247191011236  | 300                  | 0.8314606741573034 |
| 40                   | 0.8342696629213483 | 400                  | 0.8342696629213483 |
| 50                   | 0.8370786516853933 | 500                  | 0.827247191011236  |
| 60                   | 0.8356741573033708 | 600                  | 0.8258426966292135 |
| 70                   | 0.8426966292134831 | 700                  | 0.8230337078651685 |
| 80                   | 0.8342696629213483 | 800                  | 0.827247191011236  |
| 90                   | 0.8328651685393258 | 900                  | 0.8258426966292135 |
| 100                  | 0.8384831460674157 | 1000                 | 0.824438202247191  |

I used k-fold cross validation and have run the model for different number of iterations/epochs. For this dataset, however, the model gives very similar accuracy despite increasing the accuracy from 1 to even up to 1000 epochs. Although, there is a slight decrease in accuracy when epochs increase after 100.



## Code

The code is available here -

[https://colab.research.google.com/drive/1s9SDxtIGKU\\_xx0UMm7dDxxOpSURT9e48?usp=sharing](https://colab.research.google.com/drive/1s9SDxtIGKU_xx0UMm7dDxxOpSURT9e48?usp=sharing)

(Please use IITJ email id to access this notebook)