



Optimizing Neural Networks Project

Project Deliverables

You will be required to provide the following deliverables.

- A python notebook with your solution.

Instructions

Background

The sinking of the Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren’t enough lifeboats for everyone on board, resulting in the death of 1502 out of 2224 passengers and crew.

Problem Statement

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others. In this challenge, you have been tasked to build a neural network model that answers the question: “what sorts of people were more likely to survive?” using passenger data (ie name, age, gender, socio-economic class, etc). You will be required to perform feature engineering then tune your model.

You can use the following guiding notebook. [\[Link\]](#)

Hints (Feature Engineering):

1. Create a new feature title by extracting from passenger names i.e. Mrs, Miss, Mr and Master.
2. Create a new cabin number feature deck.
 - You can extract the letters before the Cabin no, which represent the deck.
3. Create a new family_size feature i.e. Parch + SibSp.
4. Create a new feature fare_per_person i.e. Fare / Family Size.

Dataset Information

You have been given two similar datasets that include passenger information like name, age, gender, socio-economic class, etc. One dataset is titled `train.csv` and the other is titled `test.csv`.

- Train.csv will contain the details of a subset of the passengers on board (891 to be exact) and importantly, will reveal whether they survived or not, also known as the “ground truth”.
- The `test.csv` dataset contains similar information but does not disclose the “ground truth” for each passenger.

The data contains features and labels:

- **Features**
 - Passenger Id, Passenger Class, Passenger Name, Passenger Gender, Passenger Age, No. of Siblings, spouse, parent and children related to passenger on-board, Passenger Ticket, Ticket's Fare, Passenger Cabin, Embarked (location on the ship).
- **Label**
 - Survived

Train Dataset = <https://bit.ly/31azYjb>

Test Dataset = <https://bit.ly/2XmmAYe>

Dataset Glossary

- survival = Survival, 0 = No, 1 = Yes
- pclass = Ticket class, 1 = 1st, 2 = 2nd, 3 = 3rd
- sex = Sex
- 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, Q = Queenstown, S = Southampton

Variable Notes

pclass: A proxy for socio-economic status (SES)

- 1st = Upper
- 2nd = Middle
- 3rd = Lower

age: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

- sibsp: The dataset defines family relations in this way:
- Sibling = brother, sister, stepbrother, stepsister

- Spouse = husband, wife (mistresses and fiancés were ignored)

parch: The dataset defines family relations in this way:

- Parent = mother, father
- Child = daughter, son, stepdaughter, stepson
- Some children travelled only with a nanny, therefore parch=0 for them.

Acknowledgements

The dataset was sourced from Kaggle [\[Link\]](#).