

Chapter 1

Introduction

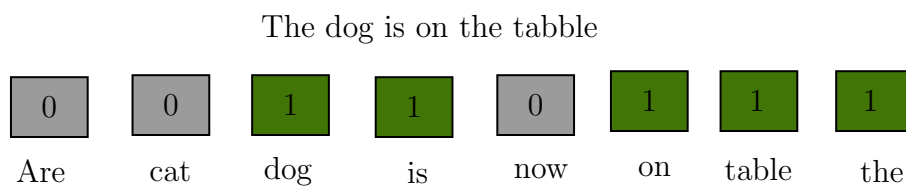
1.1 Introductionnnn

1.1.1 Feature extractionnnn from texts and image

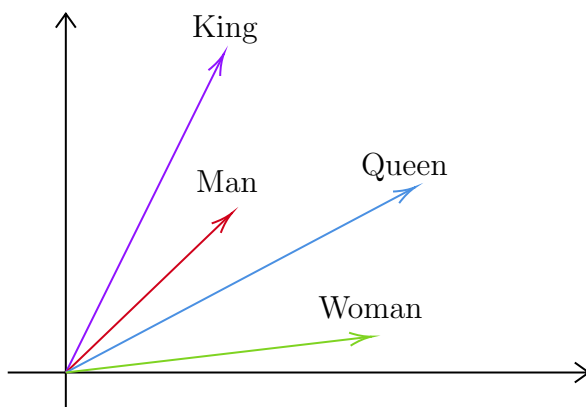
This is the titanic data sets

Text to vector

1. Bag of words:



2. Embeggins:

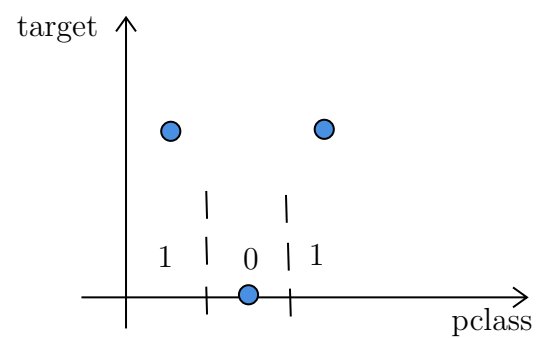
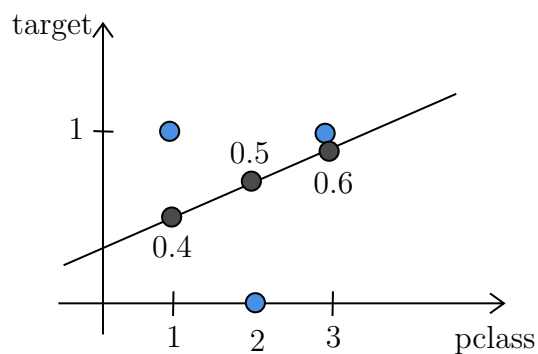


Text preprocessing

1. Lowercase
2. Lemmatization
3. Stemming
4. StopWords

1.1.2 Categorical and ordinal features**Categorical****Ordinal Features****Label encoding**

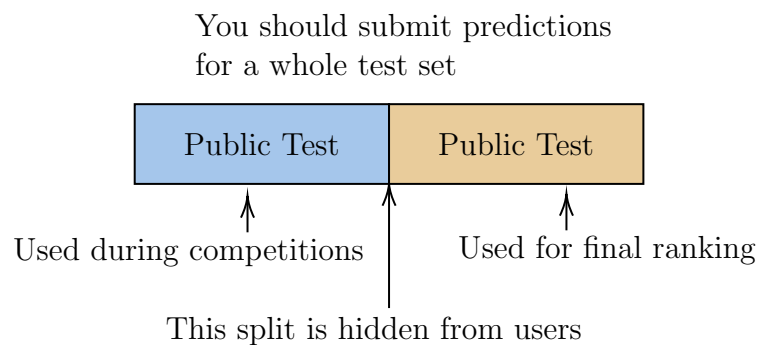
pclass	1	2	3
target	1	0	1

**Frequency encoding****1.1.3 Competition Mechanics**

1. Evaluation function Exists some evaluation functions like:
 - Accuracy
 - Logistic loss
 - AUC

- RMSE
- MAE

2. Public/Private tests



3. Sites:

- Kaggle
- Driven data
- CrowdAaqualityx
- CodaLab
- DataScienceChallenge.net
- DataScience.net
- Single-Competition sites like KDD,VizDooM

4. Conclusions:

- Main Concepts(Data, Model, Submission, Evaluation, Leaderboard)
- Competition platforms
- Reasons for participating

- 1.1.4 Datetime and Coordinates
- 1.1.5 Handling missing
- 1.1.6 Handling missing
- 1.1.7 Kaggle overview
- 1.1.8 Numeric features
- 1.1.9 Recap of main algorithms