

Your names and id

- Yuyan Fan (917229546)
- Jayoung Kim (914311526)
- Melanie Lue (917164204)

Email address

- yuyfan@ucdavis.edu
- jygtkim@ucdavis.edu
- mhlue@ucdavis.edu

Project topic

- Titanic Survival
- <https://www.kaggle.com/c/titanic/overview>

Goal

- Using machine learning, we create a model that predicts what sorts of people were more likely to survive from the Titanic.

Background

- On April 15, 1912, Titanic sank after colliding with an iceberg.
- The death of 1502 out of 2224 passengers and crew.
- While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

Algorithms/Packages

- For each in the test set, we will predict a 0 or 1 value for the variable.
- Binary classification algorithm
 - Logistic regression (create our own function)
 - Support vector machine

Challenge

- Creating own function
 - logistic regression
 - svm
- Imbalance data
 - Classification problem

Proposed work (what and how you are planning to do)

- Prepare and clean the data
 - Using passenger data (i.e. name, age, gender, socio-economic class, etc)
- Visualization
 - Histogram, box plot, scatter plot, etc.
- Explore the data
- Build models and predict
- Answer the question
 - “What sorts of people were more likely to survive?”