

Lab 10

Titanic Survival Prediction

This lab will return to the titanic dataset to predict survival of a passenger. This dataset is obtained from the `earth` package in R.

The data frame has 1046 observations on 6 variables.

- pclass passenger class, unordered factor: 1st 2nd 3rd
- survived integer: 0 or 1
- sex unordered factor: male female
- age age in years, min 0.167 max 80.0
- sibsp number of siblings or spouses aboard, integer: 0...8
- parch number of parents or children aboard, integer: 0...6

```
set.seed(11142024)
library(tidyverse)
library(earth)
data("etitanic")
titanic <- etitanic |>
  mutate(survived_factor = factor(survived))
glimpse(titanic)
```

Rows: 1,046

Columns: 7

```
$ pclass      <fct> 1st, 1st, 1st, 1st, 1st, 1st, 1st, 1st, 1st, 1st, 1st, ~
$ survived    <int> 1, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, ~
$ sex         <fct> female, male, female, male, female, male, female, male~
$ age         <dbl> 29.0000, 0.9167, 2.0000, 30.0000, 25.0000, 48.0000, 63~
$ sibsp       <int> 0, 1, 1, 1, 1, 0, 1, 0, 2, 0, 1, 1, 0, 0, 0, 0, 0, 0, ~
$ parch       <int> 0, 2, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, ~
$ survived_factor <fct> 1, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, ~
```

Question 1 (2 points)

What factors in the dataset do you think will influence whether a passenger survives? How do you expect that factors to change survival outcomes?

Question 2 (6 points)

Regardless of your response to the first question, create figures to explore survival as a function of `age`, `sex`, and `pclass`. As always, include informative titles, axes, and legends.

Question 3 (2 points)

Construct a training set and a test set. If you plan to do model tuning, also create a validation set.

Question 4 (4 points)

Use a logistic regression model to predict passenger survival. Summarize the model outcome using classification error (% of incorrect predictions on the test set).

```
#log_reg <- glm(survived_factor ~ , family = binomial(link = 'logit'), data = train_titanic)
```

Question 5 (4 points)

Use a tree-based model to predict passenger survival. Summarize the model outcome using classification error (% of incorrect predictions on the test set).

Question 6 (2 points)

Do your model outcomes match your intuition and data visualizations? Why or why not?