

RMS Titanic: Machine learning from disaster

Myrthe Boone 6E

7 februari 2019

Profielwerkstuk Wiskunde D

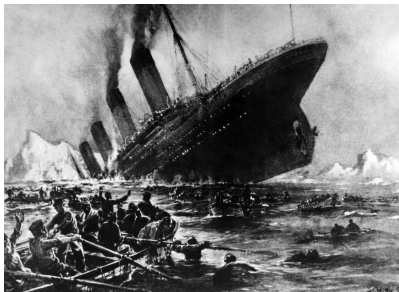


Figure: RMS Titanic

Inhoudsopgave

- ▶ Introductie
- ▶ Wat is machine learning?
- ▶ Mijn onderzoek
 - ▶ Algoritmes
 - ▶ Hoofd- en deelvragen
- ▶ Resultaten
- ▶ Conclusie
- ▶ Discussie
- ▶ Afsluiting

Waarom dit onderwerp?

- ▶ Meer weten over machine learning en programmeren in Python
- ▶ Overall om ons heen
- ▶ Organiseren en analyseren van de werkelijkheid
- ▶ TU Eindhoven

Titanic

- ▶ Voorspellen welke groep passagiers een grotere kans had om te overleven.
- ▶ 'Vrouwen en kinderen eerst' beleid
- ▶ Geluk, geslacht, leeftijd, klasse en prijs betaald voor een ticket

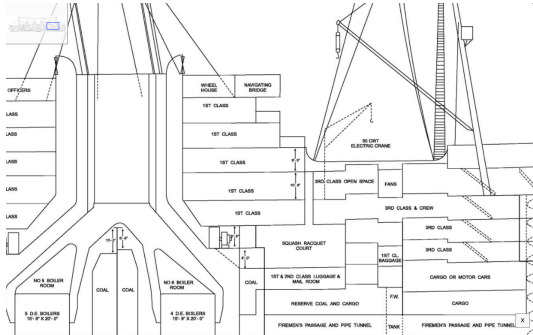


Figure: Deck Titanic

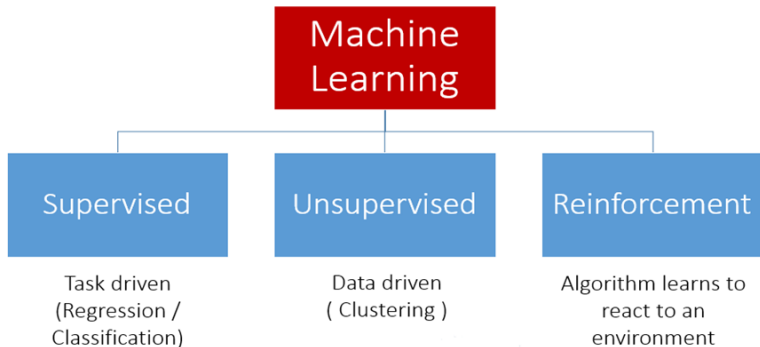
Machine learning

- ▶ Zonder specifiek geprogrammeerd te zijn voor de taken
- ▶ Voorbeeld: spam emails, Google zoekopdrachten

Soorten machine learning

- ▶ Supervised, unsupervised en reinforcement learning
- ▶ Unsupervised: Correcte labels zijn **niet** gegeven
- ▶ Supervised: Computer weet welke categorieën er zijn
- ▶ Supervised learning kan ingedeeld worden in regression en classification

Types of Machine Learning



Soorten machine learning

- ▶ Classification: categorieën
- ▶ Regression: continue variabelen

Soorten machine learning

- ▶ Titanic is supervised learning
- ▶ Binary classification
- ▶ 1 = overleefd
- ▶ 0 = niet overleefd

Werkplan

- ▶ We splitsen onze dataset in een training en een test set
- ▶ We trainen / fitten ons model op de training set
- ▶ We voorspellen op de test set
- ▶ Dus we gebruiken de gegevens van de passagiers (leeftijd, geslacht, prijs betaald voor een ticket, klasse)

Logistic Regression

- ▶ Gebaseerd op de logistische functie
- ▶ Grenswaarde $p > 0.5$, passagier heeft het overleefd

$$\sigma(y) = \frac{e^y}{1 + e^y} \quad (1)$$

- ▶ Vier variabelen dus y is in dit geval:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + b + \varepsilon_i \quad (2)$$

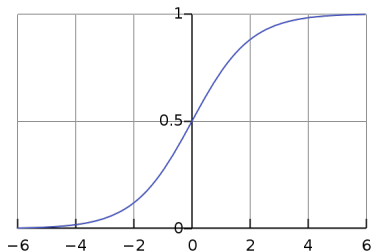


Figure: Logistische functie

Hoofd- en deelvragen

- ▶ Is het mogelijk een nauwkeurige voorspelling te maken of de passagiers aan boord van de Titanic het hebben overleefd met behulp van informatie over geslacht, klasse, leeftijd en prijs betaald voor een ticket?
 - ▶ Wat is de invloed van geslacht op de overlevingskans?
 - ▶ Wat is de invloed van klasse op de overlevingskans?
 - ▶ Wat is de invloed van leeftijd op de overlevingskans?
 - ▶ Wat is de invloed van prijs betaald voor een ticket op de overlevingskans?

Resultaten

- ▶ begonnen met plots maken, dataset ontdekken
- ▶ coëfficiënten

