



Data Advanced

Hoofdstuk 4

Machine Learning

DE HOGESCHOOL MET HET NETWERK

Hogeschool PXL – Elfde-Liniestraat 24 – B-3500 Hasselt
www.pxl.be - www.pxl.be/facebook



Inhoud

1. Inleiding
2. Historiek
3. ML problemen herkennen
4. Wanneer ML gebruiken?
5. Het ML proces
 1. Type ML - problemen
 2. Data
 3. ML - algoritme
6. Types van ML – problemen
 1. Classificatie
 2. Regressie
 3. Clustering
 4. Recommendations
7. ML – problemen uitgewerkt
8. Bronnen



1. Inleiding

- Frequent gehoorde term
- Wiskunde – statistiek?
- Machines zelfstandig taken uitvoeren
- Supervised
 - Classificatie
 - Regressie
- Unsupervised
 - Clustering
 - Recommendations



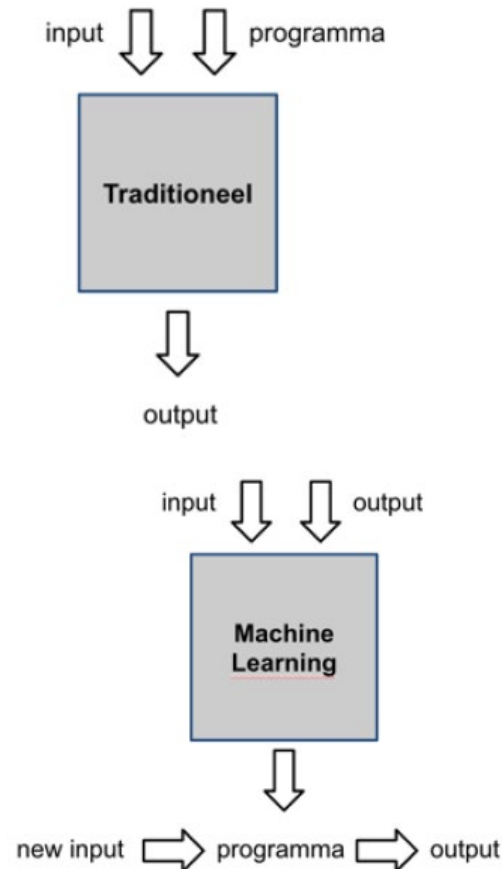
2. Historiek

- Arthur Samuel (jaren '50): schaakspel
- Frank Rosenblatt: perceptron
- Stanford University (1979): bewegende robot



3. ML problemen herkennen

- Voorbeeld: Alien
- Klassieke aanpak
- ML aanpak



3. ML problemen herkennen: vb

- Netflix
- E-mail: spam \Leftrightarrow geen spam
- Zelfrijdende auto
- Slimme thermostaat
- Gezichtsherkenning



4. Wanneer ML gebruiken?

- Google Maps
 - Statische aanpak
 - Dynamische aanpak
- Vuistregels
 - Moeilijk om regels op te stellen
 - Grote set historische data voorhanden?
 - Relaties zijn dynamisch



5. Het ML - proces

- 3 stappen
 - Welk type ML?
 - Welke data?
 - Welk algoritme?



5. Type – Data – ML Algoritme

- Type
 - Supervised
 - Classificatie: Naive Bayes, ...
 - Regressie
 - Unsupervised
 - Clustering: K – means clustering, ...
 - Recommendations
- Data
- ML algoritme



6. Types van ML - algoritmen

- Classificatie
- Regressie
- Clustering
- Recommendations



7. ML – problemen: uitgewerkt

- Classificatie
 - Classification problem statement
 - Features
 - Training
 - Testing



7. Classificatie: uitgewerkt

- Sentiment Analysis
 - Doel: tweet positief of negatief
 - Categorieën: positief / negatief
 - Classifier: Naive Bayes Algoritme
 - Training data: tweets reeds geclassificeerd
- PluralSight: How to think about Machine Learning Algorithms:
 - 4. Solving Classification Problems
 - Implementing Naive Bayes



7. Classificatie: uitgewerkt

- Reclame blokker
 - Doel: advertentie reclame of niet?
 - Categorieën: reclame / geen reclame
 - Classifier: Support Vector Machine Algoritme
 - Training data: advertenties reeds geclassificeerd
- Pluralsight: How to think about Machine Learning Algorithms:
 - 4. Solving Classification Problems
 - Implementing Support Vector Machine



7. ML – problemen: uitgewerkt

- Regressie
 - Voorspellen continue variabele
 - Verband tussen 2 variabelen
 - Problem Statement
 - Features
 - Training
 - Testing



7. Regressie: uitgewerkt

- Methode van de kleinste kwadraten
- PluralSight: How to think about Machine Learning Algorithms:
 - 6. Solving Regression Problems
 - Minimizing Error Using Stochastic Gradient Descent
 - Implementing Linear Regression in Python



7. ML – problemen: uitgewerkt

- Clustering
- Dataset
- Features
- Algoritme
- Classificatie \Leftrightarrow Clustering



7. Clustering: uitgewerkt

- Dataset: set van documenten
- Features: text → numerieke waarden
- Algoritme: K – means
 - Stap 1
 - Stap 2
 - Stap 3
- PluralSight: How to think about Machine Learning Algorithms:
 - 8. Clustering Large Data Sets into Meaningful Groups
 - Implementing K – means Clustering



7. ML – problemen: uitgewerkt

- Recommendations
- Nieuwe noden – trouw blijven
- Voorbeelden
 - Top 10 films
 - Browser geschiedenis



7. Recommendations: uitgewerkt

- Collaborative Filtering
- Latent Factor Analysis
- PluralSight: How to think about Machine Learning Algorithms:
 - 7. Recommending Relevant Product to a user
 - Implementing Alternative Least Squares to find movie Recommendations

