

Hoofdstuk 1

Supervised Learning

1.1 Supervised vs unsupervised

Machine learning kan worden opgesplitst in twee grote categorieën; *supervised learning* enerzijds en *unsupervised learning* anderzijds. De naam van beide vormen geeft eigenlijk al weg wat deze vormen precies inhouden.

Bij *unsupervised learning* hebben we voor elk datapunt in onze dataset zowel features als outputs voor handen. Er is sprake van zogenaamde 'input-outputparen'. We kunnen ons model trainen op het vinden van een verband tussen de x -waarden en de y -waarden in onze trainings-dataset. Het uiteindelijke doel is dan dat het model voor een nieuwe dataset, op basis van de features in die dataset, een output zal voorspellen.

Bij *unsupervised learning* zal de data enkel uit inputs of x -waarden bestaan en hebben we dus geen afhankelijke variabelen. Een *unsupervised learning*-model zal zelf patronen of structuren trachten te vinden in de data en vervolgens de data zelf onderverdelen in klassen. De details achter *unsupervised learning* liggen buiten de scope van dit rapport en zullen we dus achterwege laten.

1.2 Vormen van supervised learning

Supervised learning kan nog verder gesplitst worden in twee categorieën; de twee voornaamste vormen van *supervised learning* zijn regressie en classificatie.

Bij regressie zijn de outputs altijd getallen. Deze numerieke outputs kunnen eender welke waarde aannemen. We kunnen hierbij dus spreken van *kwantitatieve data*. Het getrainde model zal trachten een continu numeriek resultaat te voorspellen voor de output van een nieuw datapunt - waarop het model dus niet getraind is.

Dit staat in tegenstelling met classificatie, waar de outputs enkel een telbaar aantal waarden kunnen aannemen. We zullen in hoofdstuk ?? classificatie in meer detail toelichten.