

## **Machine Learning Basisprincipe**

Machine learning verwijst naar het proces van computers die leren van gegevens, zonder expliciet geprogrammeerde instructies. Het basisprincipe is om modellen te bouwen die patronen in data kunnen identificeren en gebruiken om voorspellingen te doen of beslissingen te nemen.

## **Supervised Learning vs. Unsupervised Learning**

Supervised Learning machine learning wordt gebruikt wanneer we gelabelde data hebben, wat betekent dat we zowel de invoer (kenmerken) als de gewenste uitvoer (labels) hebben. Het doel is om een voorspellend model te bouwen dat in staat is om correcte uitvoer te produceren, gegeven nieuwe invoer.

Unsupervised Learning is in tegenstelling tot supervised learning hebben we bij unsupervised learning ongelabelde data, wat betekent dat we alleen de invoer hebben zonder bijbehorende uitvoer. Het doel is om verborgen patronen of structuur in de data te ontdekken zonder expliciete instructies.

## **Toepassing van KMeans, Logistic Regression en Decision Trees**

KMeans is een unsupervised learning-algoritme dat wordt gebruikt voor clustering. Het verdeelt de data in k clusters, waarbij elk datapunt tot het dichtstbijzijnde clustercentrum wordt toegewezen.

Logistic Regression is een supervised learning-algoritme dat wordt gebruikt voor classificatie. Het model berekent de kans dat een bepaalde invoer tot een bepaalde klasse behoort en voorspelt de klasse met de hoogste kans.

Decision Trees is een supervised learning-algoritme dat wordt gebruikt voor classificatie en regressie. Het bouwt een boomstructuur waarin elke knoop een test op een attribuut uitvoert, en elke tak vertegenwoordigt het resultaat van de test.

## **Problemen bij het leren Over/Under-fitting**

Overfitting: Dit treedt op wanneer het model te complex is en zich te veel aanpast aan de trainingsdata, waardoor het slecht presteert op nieuwe, ongeziene data.

Underfitting: Dit treedt op wanneer het model te eenvoudig is en niet in staat is om de complexiteit van de data vast te leggen, waardoor het ook slecht presteert op nieuwe data.

## **Maatregelen tegen Overfitting/Underfitting**

Enkele maatregelen tegen overfitting zijn het gebruik van meer trainingsdata, het toepassen van regularisatie (bijv. L1- en L2-regularisatie), het gebruik van eenvoudigere modellen en het toepassen van cross-validatie.

Maatregelen tegen underfitting zijn het gebruiken van complexere modellen, het toevoegen van meer features aan de data, en het verhogen van het model complexiteit.

## **Het Testen van Modelprestaties**

Het testen van modelprestaties kan worden uitgevoerd door het model te evalueren op een aparte set data die niet is gebruikt tijdens het trainen (testset). Dit geeft een indicatie van hoe goed het model presteert op nieuwe, ongeziene data.

Door deze concepten te begrijpen en toe te passen, kun je effectieve machine learning-modellen bouwen en problemen zoals over- en underfitting herkennen en aanpakken.