

DE ENE STATISTISCHE METHODE IS DE ANDERE NIET...

Alvorens een statistische methode te gaan kiezen, dienen we goed te snappen welk probleem we willen gaan oplossen. Zonder helder probleem zijn alle antwoorden immers juist, of fout... Een fundamentele opdeling van methodes is die van supervised en unsupervised learning.

📄 DRIES VAN NIEUWENHUYSE ONDERZOEKER HOGENT



LEES DIT ARTIKEL INDIEN U WILT WETEN:

- Hoe een statistische methode te kiezen
- Wat het verschil is tussen supervised en unsupervised learning

#Skills #Controlling #Analyse

UNSUPERVISED

Beschrijvende methodes leveren geen echte uitkomst op, ze beschrijven een set van gegevens zonder voorspellingen te gaan doen. Ze laten toe van een beter begrip van de data te bekomen door meer context te bieden.

Classificatie: Wanneer we geen echte uitkomst willen berekenen en enkel een karakterisering van onze observaties wensen uit te voeren, kunnen we een clusteranalyse gebruiken. Die laat toe om gevallen in gelijke groepjes op te delen die onderling meer gelijkenissen vertonen dan in vergelijking met willekeurige andere punten. De methode berekent onderlinge afstanden en wijst punten toe aan de dichtste buur of de meest gelijkende.

Associaties: Correspondentieanalyse beeldt associaties af tussen rijen en kolommen van frequentietabellen of tellingen. Zo tonen *brand maps* verwantschappen tussen merken (in de rijen) en karakteristieken (in de kolommen) op een grafiek. Eigenschappen die op de biplot dicht bij een merk liggen, typeren dat merk. Merken die dicht bijeen liggen zijn verwant, karakteristieken die dicht bijeen liggen, zijn geassocieerd en worden samen aangetroffen bij analoge merken.

Correlaties: Principale Componenten

Analyse laat toe om te bepalen welke continue variabelen er daadwerkelijk toe doen. Deze datareductiemethode illustreert hoeveel unieke samenhangende concepten verscholen zitten in de multitude aan veranderlijken. Een biplot beeldt hier de variabelen en wanneer gewenst de observaties af in één vlak. Hoe dichter de labels van de variabelen liggen, hoe nauwer verwant. Hoe verder, hoe meer specifiek ze zijn. De techniek laat toe van het aantal KPI's te reduceren tot de nodige en voldoende.

SUPERVISED

Supervised methoden zorgen voor formules die eenmaal berekend tot een score voor een nieuwe observatie leiden. Regressietechnieken zoeken naar lineaire en niet-lineaire verbanden of proberen een kans op een gebeurtenis te voorspellen.

Lijnvormig: Lijnvormige verbanden tussen een onafhankelijke, verklarende variabele en een uitkomst of responsvariabele kunnen vastgesteld worden via lineaire regressie. Aan de hand van een vergelijking krijgen we een idee van de verandering van de onafhankelijke variabele in functie van de verandering van een of meerdere responsvariabelen. De aldus bekomen formule of vergelijking is bijzonder handig en kan in applicaties of operationele systemen worden ingebouwd. Dit noemen we het operationaliseren van analytische intelligentie.

Gebogen: Lijnvormige verbanden zijn vrij simpel, niet-lijnvormige veel moeilijker, vooral om te interpreteren en toe te lichten, niet zozeer om te berekenen. Een exponentiële groei van onze omzet bijvoorbeeld heeft een vergelijking van hogere orde die

minder intuïtief aanvoelt maar evengoed bruikbaar is wanneer de zaken extreem goed gaan. Dergelijke polynomiaalmodellen zijn ook bruikbaar voor het vaststellen van minimale en maximale impact.

Logistisch: Nog complexer is logistische regressie. Door een transformatie van de cijfers bekomen we een S-vormige curve die schommelt tussen 0 en 1, ideaal dus om de kans te berekenen dat een gebeurtenis zich zal voordoen. Dergelijke modellen berekenen de probabiliteit dat iemand een bepaald product zou aanschaffen bijvoorbeeld op basis van de huidige gekochte producten.

DE VRAAG IS BELANGRIJKER DAN HET ANTWOORD...

Voor zowat elk soort vraag werd ooit wel een supervised of een unsupervised methode ontworpen die ons verstaanbare output zal opleveren die bruikbaar wordt binnen de organisatie. Zo kunnen we bepalen bij welke soorten klanten we wat en hoeveel zouden kunnen verkopen. Daarmee kunnen we onze bedrijfsvoering substantieel verbeteren. De juiste vraag stellen is in deze een cruciale start, een relevante methode kiezen allesbepalend om tot zinvolle antwoorden en dito besluiten te komen. Consistente resultaten door het gebruik van meerdere complementaire methoden laten toe om methodologische fouten tijdig te detecteren. Om het even welke methode gebruikt wordt, de resultaten zouden moeten in lijn liggen van elkaar. Het probleem vanuit verschillende invalshoeken bekijken zal voor een genuanceerd beeld zorgen dat zakelijk de boel vooruithelpt. En daar gaat het tenslotte om.