# [Discriminantanalyse](https://datascience2.netlify.app/ch6_lda/1_inleiding)

## Enkele begrippen

Univariate statistiek: 1 variabele, bivariate statistiek: 2 variabelen nodig

Voorbeeld bivariate: regressieanalyse met onafhankelijke (x) en afhankelijke (y)

Wanneer doelgroepen elkaar niet overlappen; waarneming behoort tot 1 groep 🡺 **wederzijds uitsluitende groepen** (bv. Man/vrouw)

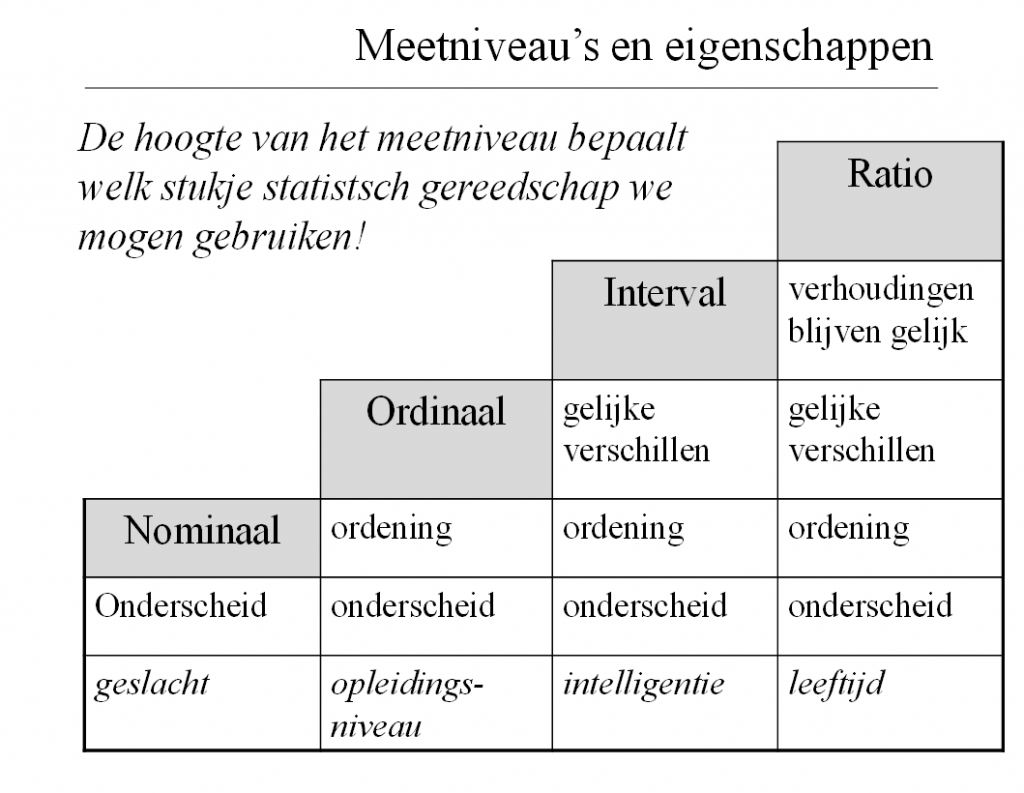
## Wat is discriminantanalyse?

Discriminantanalyse behoort tot multivariate statistiek: bepalen tot welke gegeven groepen de waarnemingen best thuishoren.

Voorbeeld met bank: krediet of niet-kredietwaardig

## Afhankelijke en onafhankelijke variabelen

1. Afhankelijke variabele bij discriminantanalyse is de categorie: krediet- of niet-kredietwaardig.
2. Onafhankelijke variabelen zijn gegevens die worden gebruikt om tot de categorie te komen. Bv. Stabiliteit, financiële draagkracht en leeftijd



## Karakteristieken

Kan gebruikt worden voor twee doeleinden

1. Verschillen tussen wederzijds uitsluitende groepen weergeven door afhankelijke variabele (categorie) door patronen te vinden in de waarde van onafhankelijke: **descriptieve discriminantanalyse**Relatieve belangrijkheid van kenmerken/eigenschappen van verschillende groepen onderscheiden/bepalen. Hoe beter de differentiatie, hoe hoger het gewicht.

Het resultaat is de discriminantfunctie. Hierbij wordt de waarde voor elk van de discriminanten vermenigvuldigd met een gewicht bij elkaar opgeteld (lineaire descriptieve discriminant analyse). Kan je toepassen op waarneming. Neemt enkel waarden die overeenstemmen met discriminanten en vermenigvuldigd deze met de overeenstemmende gewichten 🡺 **discriminantfunctiescore**

*En word uiteindelijk gebruikt voor predictieve discriminantanalyse*

1. Bepalen tot welke groep een nieuwe waarneming behoort: **predictieve discriminantanalyse**

## Veronderstellingen

1. Er is geen afhankelijkheid tussen de onafhankelijke variabelen
2. Tussen elk paar van onafhankelijke variabelen is er lineairiteit
3. Geen co-lineairiteit tussen onafhankelijke variabelen
4. Multivariate normaliteit: komt overeen met normale verdeling
5. Homogeniteit van de variantie-covariantiematrices

## Waar wordt discriminantanalyse gebruikt?

Word gebruikt bij patroonherkenning en in machine learning om een lineaire combinatie van kenmerken te vinden die twee of meer klassen van objecten of gebeurtenissen van elkaar scheidt. Voorspellen van bv. Faillissementen. Aantal kenmerken beperken tot beheersbaar aantal.

# Evaluatiemetrieken

# Neurale netwerken

Stap 0: Laad dataset op en inspecteer gegevens

Stap 2: Voer de nodige data management manipulaties om de data verwerking klaar te maken

Stap 3: Normaliseer de data (alleen indien nodig en indien normalisatie geen deel uit maakt van het gekozen ANN-model)

Stap 4: Indien vereist, splits de dataset op in een training- en test dataset

Stap 5: Construeer het ANN-model

Stap 6: Train het ANN-model

Stap 7: Evalueer de kwaliteit van het ANN-model

Stap 8: Pas het ANN-model toe op een nieuwe dataset

**Opmerking: Afhankelijk van het project kunnen bepaalde stappen worden overslaan**

**Afbeelding met tekst, schermopname, Lettertype, nummer

Automatisch gegenereerde beschrijving**

**Afbeelding met tekst, Lettertype, schermopname, lijn

Automatisch gegenereerde beschrijving**

**Afbeelding met tekst, schermopname, Lettertype

Automatisch gegenereerde beschrijving**

**Afbeelding met tekst, schermopname, Lettertype

Automatisch gegenereerde beschrijving**

# Metaheuristieken

<https://github.com/EliasDeHondt/DataScience2>

### Question 5: Quiz

Let's answer the quiz questions:

**a. Have above-average chromosomes always a better offspring?**

No, above-average chromosomes do not always guarantee better offspring. Genetic algorithms use crossover and mutation to introduce variability and explore the search space. Sometimes, combining two above-average chromosomes can result in below-average offspring due to the interaction between their genes. Additionally, the mutation process can introduce beneficial or detrimental changes randomly.

**b. Can you explore the entire search space with crossover? Are there limitations to this exploration?**

Crossover alone cannot explore the entire search space. It combines existing genetic material from the parent chromosomes but does not introduce new genetic material. Therefore, crossover is limited to the genetic diversity present in the initial population. Mutation is necessary to introduce new genetic variations and help the algorithm explore the entire search space. Without mutation, the algorithm may converge prematurely to local optima.

**c. Why does the mutation probability play a significant role?**

The mutation probability plays a significant role because it determines how often new genetic material is introduced into the population. If the mutation probability is too low, the algorithm may not explore enough of the search space, leading to premature convergence. If the mutation probability is too high, the algorithm may become too random and lose the beneficial traits it has accumulated. A balanced mutation probability helps maintain diversity in the population and allows the algorithm to explore new areas of the search space while preserving good solutions.

**d. Why do you start simulated annealing with a randomized vector?**

Starting simulated annealing with a randomized vector ensures that the algorithm begins its search from a diverse point in the solution space. This helps prevent bias towards any particular region and allows the algorithm to explore a wider area initially. The randomness increases the likelihood of finding a global optimum rather than getting stuck in a local optimum. As the temperature decreases, the algorithm focuses on refining the solution by exploring the neighborhood of the current best solution.

**e. Why must the temperature decrease in simulated annealing?**

The temperature in simulated annealing must decrease to gradually reduce the algorithm's acceptance of worse solutions. At high temperatures, the algorithm is more likely to accept worse solutions to escape local optima and explore the search space. As the temperature decreases, the algorithm becomes more selective and focuses on refining the current solution, converging to a local or global optimum. This process mimics the physical annealing process, where slow cooling allows the system to settle into a stable state with minimal energy.

These answers provide a comprehensive understanding of the concepts behind genetic algorithms and simulated annealing, addressing the quiz questions effectively.