

Eindvergadering ML sessie 5

Team Euler: *Daan, Marie, Zeineb, Florian, Vincent, Jasper, Lasha, Younes*

Vrijdag 10 november 2023

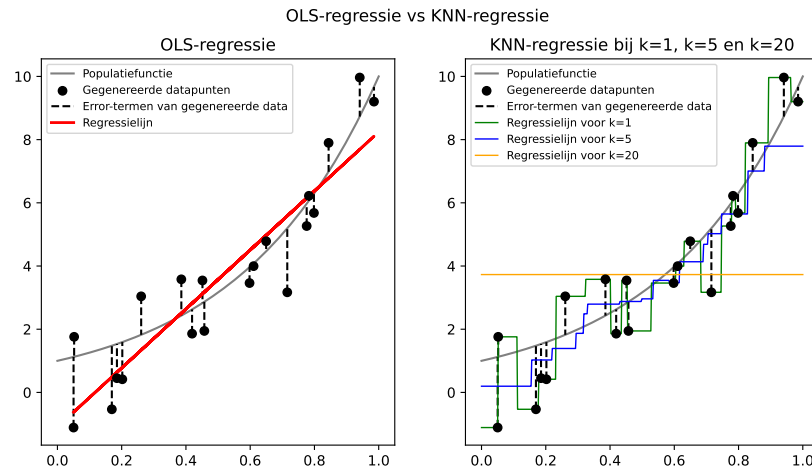
Reflectie

Jasper en Younes hebben teksten gelezen i.v.m. onze extra toepassing, Support Vector Machines. Ze zijn daarnaast ook begonnen aan de uitleg en bijhorende illustraties.

Lasha, Zeineb, Marie, Florian en Daan zijn begonnen aan de implementaties en zullen afwerken waar ze mee bezig waren. De status van het tabblad *Implementatie* in de KanBan is te zien in figuur 2.

Simulatie 1

Vóór de aanvang van de sessie was simulatie 1 reeds in orde gebracht. In deze simulatie hebben we een steekproef $(x_i, 10^{x_i} + \epsilon_i)$ van 20 punten met een standaardnormaal verdeeld residu en een uniform verdeelde regressor tussen 0 en 1. We stelden de steekproef grafisch voor, voegden de populatiefunctie toe, de regressielijn en de KNN-modellen voor $K = 1$, $K = 5$ en $K = 20$:



Figuur 1: OLS-regressie vergeleken met KNN-regressie voor verschillende waarden van K .

De OLS-regressielijn is lineair, wat ervoor zorgt dat de populatiefunctie - die niet lineair is - niet super goed benaderd wordt.

Bij de (groene) KNN-regressielijn voor $K = 1$ is er duidelijk sprake van *overfitting*: de regressielijn volgt elk datapunt perfect en wordt dus ook heel sterk beïnvloed door uitschieters. De (blauwe) regressielijn voor $K = 5$ heeft een lagere bias, aangezien deze regressielijn zich op de 5 dichtstbijzijnde burens baseert en dus veel minder gevoelig is voor uitschieters in de gegenereerde dataset. Wanneer $K = 20$, is de KNN-regressielijn een rechte waarvan de y-waarde gelijk is aan de gemiddelde y-waarde van alle datapunten, aangezien het aantal datapunten ook gelijk is aan 20. We besluiten dus dat de regressielijn voor $K = 5$ de beste schatter is.

Te doen (type+naam)	Bezig	Afgewerkt	Bestand	Finaal
Procedure xsample(n,a,b)		Vincent	procedures.py	OK
Voorbeelden xsample(n,a,b)		Vincent	voorbeelden.ipynb	OK
Procedure ysample(x,f,e)		Vincent	procedures.py	OK
Voorbeelden ysample(x,f,e)		Vincent	voorbeelden.ipynb	OK
Procedure ols(y,x)		Vincent	procedures.py	OK
Voorbeelden ols(y,x)		Vincent	voorbeelden.ipynb	OK
Procedure knn(x0,y,x,k)		Vincent	procedures.py	OK
Voorbeelden knn(x0,y,x,k)		Vincent	voorbeelden.ipynb	OK
Simulatie 1		Vincent	simulaties.ipynb, illustraties/simulatie1.pdf	OK
Simulatie 2	Daan			bezig
Simulatie 3	Marie			bezig
Procedure mls(y,X)	Vincent			bezig
Voorbeelden mls(y,X)				
simulatie 5	florian			bezig
Simulatie 6	Lasha			bezig
Simulatie 7				
Simulatie 8	Zeineb			bezig
Procedure mls(y,X)	Lasha			nakijken

Figuur 2: De status van het tabblad *Implementatie* in de KanBan aan het einde van sessie 5

Vooruitblik

De teamleden die startten met de implementatie van de basismethoden, zullen naar volgende week toe proberen afwerken waarmee ze bezig waren. Florian zal bovendien ook nog een oefening afwerken.