

Examen Machinaal Leren

Eerste zittijd
19 januari 2022

Deel B

Voornaam: _____


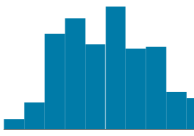
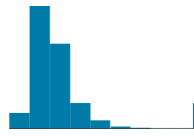
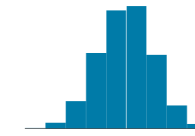

Familienaam: _____

DRAAI DEZE PAGINA PAS OM NA HET SEIN VAN DE BEGELEIDERS

- Dit examen is gesloten boek. Er zijn geen hulpmiddelen toegelaten.
- Haal deze bundel niet uit elkaar.
- Noteer je antwoorden in deze bundel. Bij sommige vragen moet je een uitdrukking of waarde invullen in de daartoe voorziene ruimte én je antwoord motiveren. Vergeet deze motivatie niet. Zonder motivatie krijg je geen punten, zelfs al is je antwoord correct.
- Schrijf leesbaar.
- Achteraan deze bundel vind je een aantal lege bladen. Duid op elk blad goed aan of het over een kladblad of antwoordblad gaat.

1. Data preprocessing [7 punten]

Onderstaande tabel vat een dataset voor “predictive maintenance” samen. Het doel is om aan de hand van machine learning te voorspellen wanneer een elektromotor zal falen aan de hand van observaties zoals de temperatuur, de snelheid en het koppel (torque). We beschouwen dit probleem als een classificatieprobleem.

Machine type	Temperature	Speed	Torque	Failed
				
A: 5% B: 80% Other: 15%	min: 10 max: 150 mean: 60	min: 0 max: 2000 mean: 100	min: -100 max: 100 mean: 10	No failure: 95% Failure: 5%

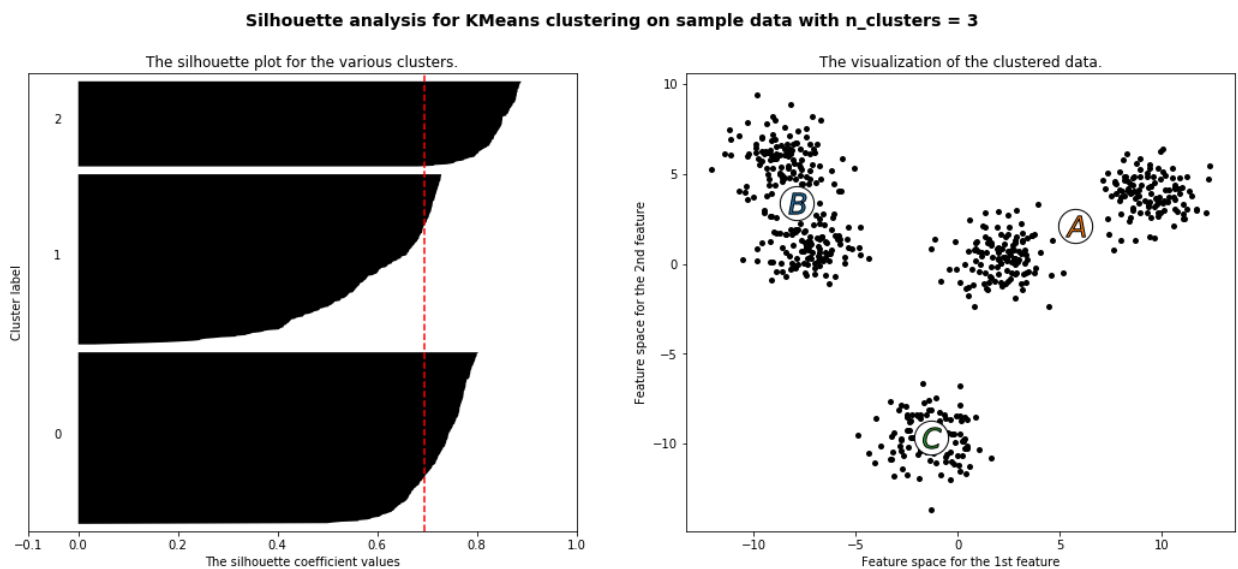
- (a) Geef 3 mogelijke problemen die je zou verwachten bij deze dataset. Verklaar kort waarom dit een probleem is en wat je zou doen om dit op te lossen.

Antwoord:

- (b) Welke metriek zou je gebruiken om je model te evalueren ? Verantwoord kort je keuze.
Antwoord:

2. Clustering met K-means [4 punten]

Onderstaande figuur toont rechts een dataset met twee features. De letters A,B en C duiden de drie gevonden clusters aan na het uitvoeren van het K-means algoritme met $K = 3$. In de linker figuur wordt de overeenkomstige silhouette plot getoond.



- (a) Geef voor elk van de cluster labels uit de silhouette plot (0,1 en 2) aan met welke cluster (A, B of C) ze overeenkomen.

Antwoord:

(b) Hoe wordt deze silhouette plot opgebouwd? Wat wordt er juist voorgesteld ?

Antwoord:

3. **SVM [4 punten]**

De (vereenvoudigde) signatuur van de Scikit-learn SVM classifier (SVC) is als volgt:

```
class sklearn.svm.SVC(C=1.0, kernel='poly ')
```

- (a) Waarvoor dient de C parameter ? Hoe kies je een waarde voor deze parameter ?

Antwoord:

- (b) Waarvoor dient de “kernel” parameter ? Geef een andere mogelijke waarde voor deze parameter.

Antwoord:

4. **Gaussian Mixture models [5 punten]**

Wat is een Gaussian Mixture model ? Zorg dat je in je uitleg de termen “Gaussian” en “Mixture” duidelijk verklaart. Geef ook aan waarvoor ze kunnen gebruikt worden en wat hun parameters zijn.

Antwoord: