**Academiejaar 2016-2017, 3 februari 2017, 8h30**

**Machine Learning**

---

Schrijf je naam duidelijk op elk antwoordenblad, alsook op het vragenblad. Beantwoord elke vraag apart, en vermeld duidelijk het nummer van elke vraag bij je antwoord. Als je een vraag niet beantwoordt, schrijf dan duidelijk het nummer van de vraag met hierbij "Geen antwoord".

---

1. **Beslissingsbomen (2pt)**
   Beslissingsbomen zijn ervoor gekend dat ze vaak leiden tot overfitting. Welke technieken zou je kunnen gebruiken om overfitting bij beslissingsbomen te reduceren. Leg elk van de technieken kort uit.

2. **Support Vector Machines (2pt)**
   Gegeven een classificatieprobleem waar de doelstelling is om, gegeven een training set, de testpunten zo goed mogelijk te voorspellen. Voorts weten we ook dat de data van sensoren afkomstig is waar geregeld ruis op zit, zodanig dat het model dient te vermijden om te veel belang aan specifieke datapunten te hechten. Om dit probleem op te lossen zullen we een SVM model met een kwadratische kernel gebruiken (polynomiale kernelfuctie van graad 2). De dataset is gegeven in de Figuur 1, en de slack penalty C van het SVM model zal dus bepalen hoe de beslissingsfunctie eruit ziet.

   (a) Waar zou de beslissingsgrens liggen voor zeer grote waarden van C (bijvoorbeeld wanneer C naar oneindig gaat) ? Teken je antwoord op de figuur hieronder. Motiveer waarom.

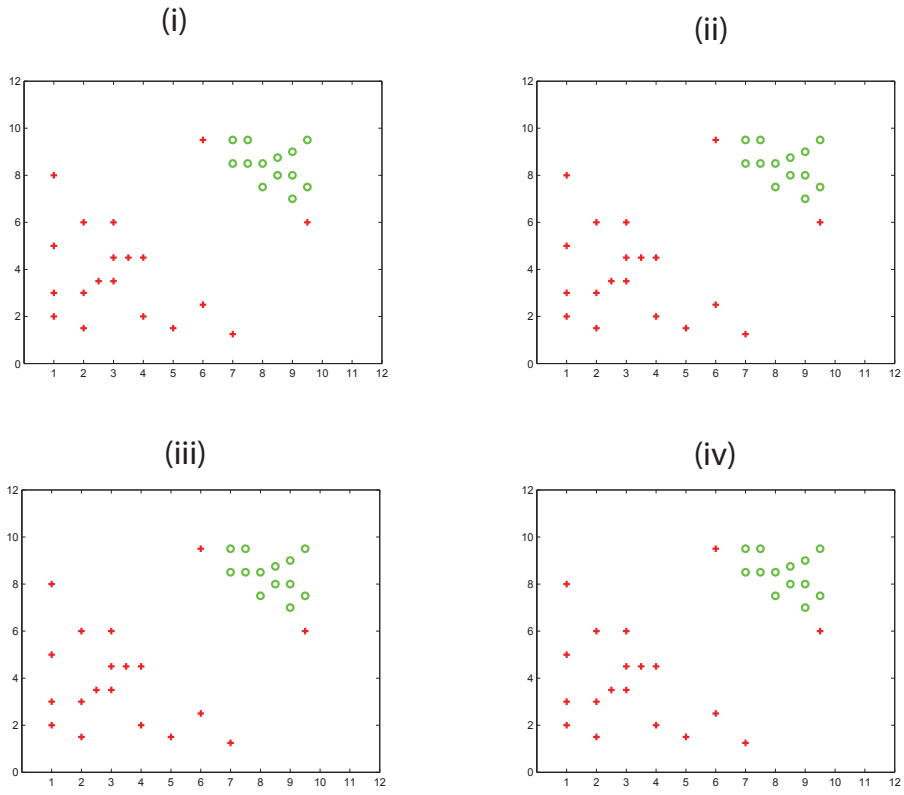
   (b) Duid op de figuur aan waar de beslissingsgrens zou liggen wanneer C tot 0 nadert. Teken je antwoord op de figuur en motiveer waarom.

   (c) Teken een datapunt dat de beslissingsgrens voor zeer grote waarden van C niet verandert. Motiveer waarom.

   (d) Teken een datapunt dat de beslissingsgrens voor zeer grote waarden van C drastisch verandert. Motiveer waarom.

3. **Neurale netwerken**

   - Bewijs de convergentie van de perceptron leerregel (2pt)
   - Een enthousiaste dokter van het UZ Gent heeft over machine learning gehoord en vraagt om jouw hulp. Hij heeft gedurende 2012 data van patienten verzameld en heeft voor elke patient een aantal features opgemeten (bvb. temperatuur, lengte, gewicht,...). Hij wil weten of hij een model kan opstellen dat voor nieuwe patienten kan beslissen of ze diabetes, een hartziekte of Alzheimer hebben (een patient kan meerdere ziekten tegelijk hebben).
     Je beslist om hiervoor een neuraal netwerk te gebruiken, maar wordt geconfronteerd met verschillende opties: ofwel een apart netwerk trainen voor elk van de ziekten, ofwel een enkel neuraal netwerk met 1 output neuron per ziekte, maar met een gedeelde verborgen laag. Welke methode zou je prefereren ? Motiveer je keuze. (1pt)




Figuur 1: SVM dataset.

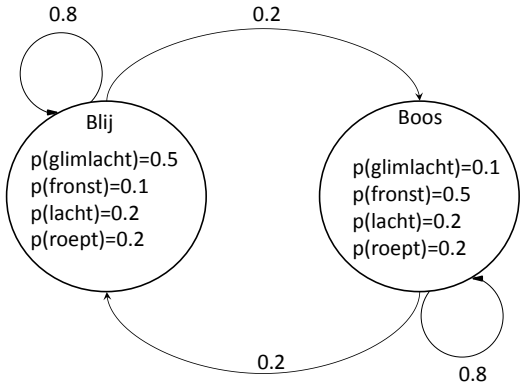
4. **Clustering**

   (a) Beschrijf de component "cooperation" bij Self Organizing Maps, en leg uit welk effect dit heeft op de clustering. (1pt)

   (b) Beschrijf twee manieren om de kwaliteit van de output van een clusteringsmethode te meten. Leg vervolgens uit hoe een statistisch raamwerk gebruikt kan worden om te testen of het resultaat van de clustering significant verschilt van random data. (2pt)

   (c) Geef één voordeel van hierarchische clustering ten opzichte van K-means clustering, en één voordeel van K-means clustering ten opzichte van hierarchische clustering. (1pt)

5. **Hidden Markov Models (3pt)** Jimmy leidt een simpel leven. Sommige dagen is hij Boos en andere dagen is hij Blij. Maar hij verbergt zijn emotionele toestand, zodat het enige wat je kan observeren is of hij glimlacht, zijn wenkbrouwen fronst, lacht of roept. We starten op dag 1 in de toestand "Blij" en veronderstellen dat er één transitie per dag is.




   Definities: $q_t$: toestand op dag $t$, $O_t$: observatie op dag $t$

   Bereken de volgende probabiliteiten:

   - $p(q_2 = \text{Blij})$
   - $p(O_2 = \text{fronst})$
   - $p(q_2 = \text{Blij} | O_2 = \text{fronst})$
   - $p(O_{100} = \text{roept})$

   We veronderstellen dat $O_1 = \text{fronst}$, $O_2 = \text{fronst}$, $O_3 = \text{fronst}$, $O_4 = \text{fronst}$ en $O_5 = \text{fronst}$. Wat is de meest waarschijnlijke bijhorende sequentie van toestanden ?

6. **Feature en prototype selection**

   (a) Beschouw een classificatiemodel in een situatie met 1000 features in totaal. 50 van deze bevatten relevante informatie met betrekking tot het klasselabel. Nog 50 andere features zijn directe kopieen van de eerste features, en de resterende 900 features zijn random gegenereerd.
   We nemen aan dat er genoeg data is om betrouwbaar te kunnen zeggen hoe nuttig bepaalde features zijn, en dat de feature selectiemethoden optimale thresholds gebruiken. Hoeveel features zullen er geselecteerd worden door:

i. een simpele filter feature selectiemethode zoals information gain

ii. een wrappergebaseerde zoekmethode

iii. een embedded methode die één enkele beslissingsboom gebruikt

iv. een embedded methode op basis van Random Forests.

Motiveer waarom. (2pt)

(b) ENN algoritme (4 pt)

- Formuleer het ENN algoritme en zijn parameters.
- Plaats deze methode in de taxonomie van PS methoden. Verklaar.
- Is het mogelijk dat deze methode alle elementen in een dataset verwijdert? Zo ja, geef een voorbeeld. Zo neen, bewijs.
- Het All-KNN algoritme (Algoritme 1) is gerelateerd aan ENN. Plaats ook deze methode in de taxonomie van PS methoden. Bespreek de verschillen met ENN en vergelijk de verwachte performantie van de twee methoden op basis van gepaste criteria.

**Algorithm 1** All-KNN

**Input:** Dataset $T$, waarde $k_{max}$
**Output:** Subset $S$

1: $M \leftarrow \emptyset$
2: **for** $k = 1, ..., k_{max}$ **do**
3:  **for each** $x \in T$ **do**
4:  Classificeer $x$ met het $k$-dichtste buur algoritme o.b.v. $T \setminus \{x\}$
5:  **if** $x$ verkeerd geclassificeerd **then**
6:  $M \leftarrow M \cup \{x\}$
7: $S \leftarrow T \setminus M$