Exam Machine Learning for the Quantified Self

12. 07. 2018

12:00 - 14:45

NOTES:

1. YOUR NAME MUST BE WRITTEN ON EACH SHEET IN CAPITALS.
2. Answer the questions in Dutch or English.
3. Points to be collected: 90, free gift: 10 points, maximum total: 100 points.
4. Grade: total number of points divided by 10.
5. This is a closed book exam (no materials are allowed).
6. You are allowed to use a SIMPLE calculator.

QUESTIONS

1. **Introduction and Removal of Sensory Noise (20 pt)**

   Cristiano is a professional athlete and normally has a very positive attitude and a lot of self-confidence. However due to some setbacks in his professional life he has ended up in a depressed state. His therapist advised him to use a dedicated app to manage his condition. This app uses the sensor information of the phone to track activities and based on the observed sensory information provides suggestions on activities that might be best for Cristiano. To collect data on the impact of activities, the app asks for self-rating every now and then (e.g. "Cristiano, what is your mood at the moment?").

   (a) **(3 pt)** Provide the definition of the quantified self we have discussed in the book and argue whether Cristiano complies to this definition.

   (b) **(3 pt)** Identify a supervised learning task for the case of Cristiano.

   (c) **(3 pt)** What would be an appropriate step size for the dataset to create a machine learning dataset for the supervised machine learning task you have just identified? Argue why.

(d) **(4 pt)** We want to apply an outlier detection on our data. We are in doubt between using Chauvenet's criterion and a Simple Distance-Based Approach. List two advantages of using the Simple Distance-Based Approach over Chauvenet's criterion.

(e) **(7 pt)** Consider Figure 1, showing a part of the sensor data of Cristiano. We want to use a lowpass filter to filter out the high frequency noise we observe in the signal. Argue what would be a suitable frequency and show what figure results after applying the lowpass filter.

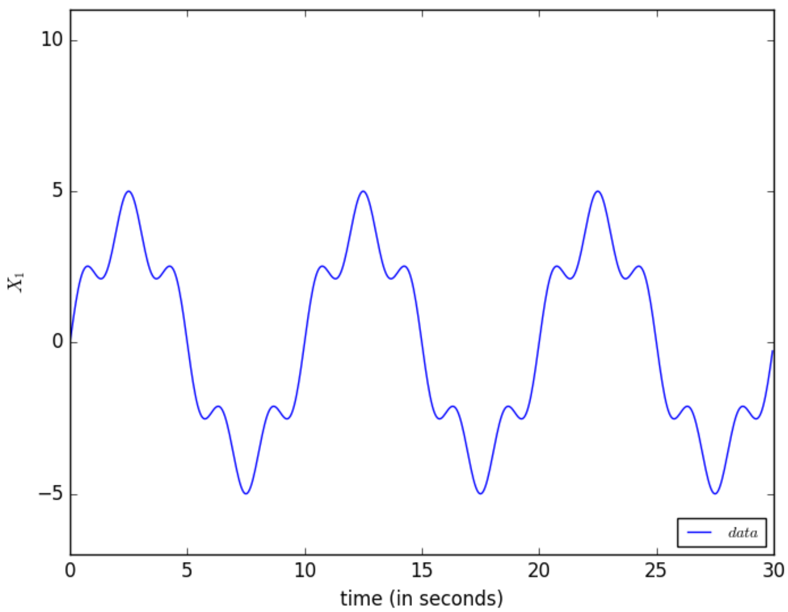

Figure 1: Example dataset

2. **Feature Engineering (15 pt)**

(a) **(3 pt)** Explain the difference between the frequency and the time domain.

(b) **(3 pt)** Imagine we know that only the most dominant frequency is relevant for making predictions in a dataset. Which of the three aggregation functions for the frequency domain would be best to apply for this case? Argue why.

Table 1: Example dataset 1

Time point	*Respiration*	*Tired*
0	30	no
1	20	no
2	50	yes
3	30	yes
4	30	no

(c) **(5 pt)** Consider Table 1. Extend the table with a feature in the time domain for the respiration attribute in which you take the mean with a value of $\lambda = 1$. Explain how you came to your answer.

(d) **(4 pt)** Provide and explain the NLP pipeline that has been explained during the lecture which we use before we can start identifying attributes in text data.

3. **Clustering (20 pt)**

Consider the data shown in Table 2, involving multiple people.

(a) **(5 pt)** Given the table in the data, we want to apply a feature based distance metric to compute the distance between the two time-series. Assume we use the mean as feature. Compute the distance between the time series of person 1 and person 2.

Table 2: Example dataset 2

Person 1	
Time point	*Acc. X*
1	8
2	-8
3	8
4	-8

Person 2	
Time point	*Acc. X*
1	-11
2	12
3	-11
4	12

(b) **(6 pt)** Compute the distance between the two time series using the cross correlation coefficient. Show how you came to your answer. In this explanation, also show what the optimal shift τ is.

(c) **(4 pt)** Explain the difference between k-means clustering and agglomerative clustering.

(d) **(5 pt)** Explain the subspace clustering algorithm in words. What is the advantage of the clustering method compared to the other clustering approaches that have been discussed during the course?

4. **Predictive Modeling (15 pt)**

Imagine that we have collected datasets around three people, identified by qs_1, qs_2, and qs_3. We identify their data instances as $(x_{qs_1,1}, \ldots, x_{qs_1,N_{qs_1}})$ for qs_1 (where N_{qs_1} is the number of data points available for qs_1), and similarly we have $(x_{qs_2,1}, \ldots, x_{qs_2,N_{qs_2}})$ and $(x_{qs_3,1}, \ldots, x_{qs_3,N_{qs_3}})$.

(a) **(4 pt)** Imagine we want to apply predictive modeling on a population level for unseen data of known users, and we assume a temporal ordering in the dataset. Specify what data would go into our training set and what data would go into our test set (you can assume a 60/40 split). Argue how you came to your answer.

(b) **(4 pt)** Explain the concept of PAC learnability and how it relates to the VC dimension.

(c) **(3 pt)** We want to tune the parameters of our predictive model, explain the best way to setup this process to perform parameter tuning. Include the parts of the data you use in your explanation.

(d) **(4 pt)** When we do parameter tuning on recurrent neural networks and echo state networks, which of the two networks would you expect to require the largest number of neurons? Argue why.

5. **Reinforcement Learning (20 pt)**

We are going to focus on a reinforcement learning case. We focus on an app we have developed to motivate students to study better and make them pass their exams. We assume a student can be in five states: *partying*, *studying*, *passed exam*, *failed exam*, and *just woke up*. The initial state of the student is assumed to be *just woke up*. We assume that we can take two actions: (1) send a message to take it easy and take time for *reflection*, or (2) send a message to *start studying*. When we send a message to start studying when the student has just woken up, the student will always go to the studying state. If we however send a message to reflect when the student just woke up he will certainly go to a partying state. Independent of the action performed in the partying state, the student will always end up in the failed the exam state. In the studying state however, when we send a message to start studying again, the student will stress out too much and ends up failing the exam. When a message is sent to reflect in the studying state the student will pass the exam. The rewards per state are shown in Table 3.

Table 3: Rewards

State	*Reward*
partying	5
studying	10
passed exam	100
failed exam	0
just woke up	5

(a) **(5 pt)** Provide a graphical representation of the Markov Decision Process that has just been described.

(b) **(4 pt)** Explain the difference between the two reinforcement learning algorithms *Q-learning* and *SARSA*.

(c) **(5 pt)** Compute the final Q-values for the state-action pairs for the case that has been described provided that the Q-learning algorithm is used. Show how you come to your answer.

(d) **(3 pt)** Explain the concept of eligibility traces.

(e) **(3 pt)** Normally, we assume all state-action pairs are stored in a table. Explain why this is not always feasible and what would be an appropriate approach to remedy this.