Exam Machine Learning for the Quantified Self
21. 06. 2017
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 (15 pt)**

   Let us consider Harry. Harry is a typical Apple lover and he buys every single device Apple decides to put on the market. As a result, Harry wears an Apple Watch all the time and also carries his phone around with him at every imaginable moment. While he wasn't such a fan of tracking his health state at first, since he suffered from a stroke he decided to start living a healthier life and he is using his Apple products to assist him. His watch for example, provides him with a goal of an amount of physical activity per day and measures how close he is to his target for the day. He closely monitors this information and tries to reach the goals set for him.

   (a) **(3 pt)** Would Harry adhere to our definition of the Quantified Self? Explain why (not).

   (b) **(4 pt)** Independent of whether you call Harry a Quantified Self, how would you categorize him in Five-Factor-Framework of Self-Tracking Motivations? Explain how you came to your answer.

   (c) **(4 pt)** Identify two machine learning tasks that can potentially result in insights to better assist Harry.

   (d) **(4 pt)** For each of the two machine learning tasks, argue what would be a proper step size (i.e. granularity for the dataset) for this task, assuming that we collect a temporal dataset.

2. **Feature Engineering (20 pt)**

   Consider the dataset shown in Figure 1. The figure shows time on the x-axis and the value for attribute X_1 on the y-axis. This dataset shows some periodic behavior we want to exploit.

   (a) **(5 pt)** Describe on a conceptual level what a Fourier transformation is.

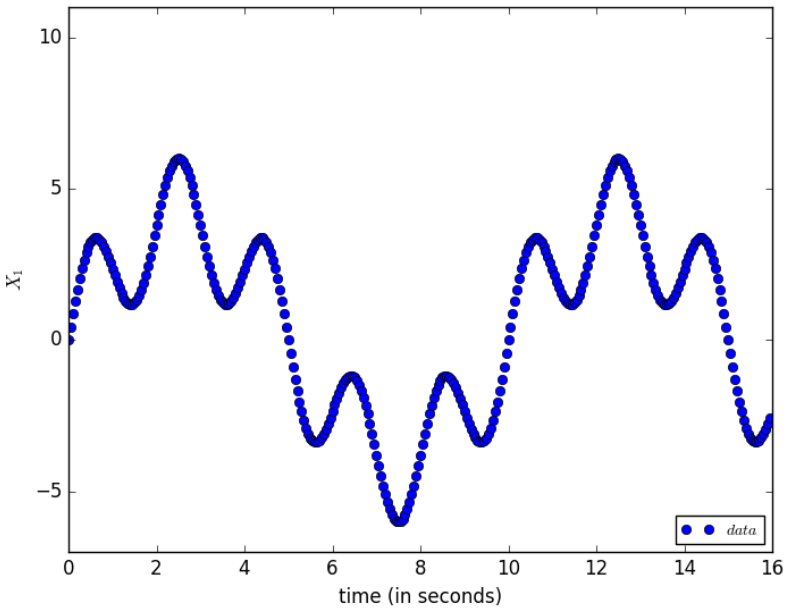

Figure 1: Example temporal dataset

(b) **(6 pt)** If we would apply a Fourier transformation to the data shown above, which frequency/what frequencies would you expect to have a high amplitude. Explain why.

(c) **(3 pt)** List three features that summarize the frequencies and amplitudes into a single number and explain how they summarize the values.

(d) **(6 pt)** In addition to the frequency domain, what other domain is available to derive useful features from temporal data? Explain how features can be derived in that domain.

3. **Clustering (20 pt)**

Consider the data shown in Figure 2. We see two attributes, X_1 and X_2, both having values in the range [0,1].

(a) **(3 pt)** Explain the purpose of subspace clustering and provide a textual description of the algorithm.

(b) **(2 pt)** Why is subspace clustering appropriate for the setting of the Quantified Self?

(c) **(6 pt)** Imagine we select a unit to be dense if it contains at least 5 data points, and we create 5 distinct intervals per attribute. Calculate what units would be found with subspace clustering given the dataset shown in Figure 2. Show your calculations.

(d) **(6 pt)** Name and explain two raw-based person level distance metrics that have been treated during the lectures.

(e) **(3 pt)** Would k-means clustering be suitable for person level clustering as well? Argue why (not).

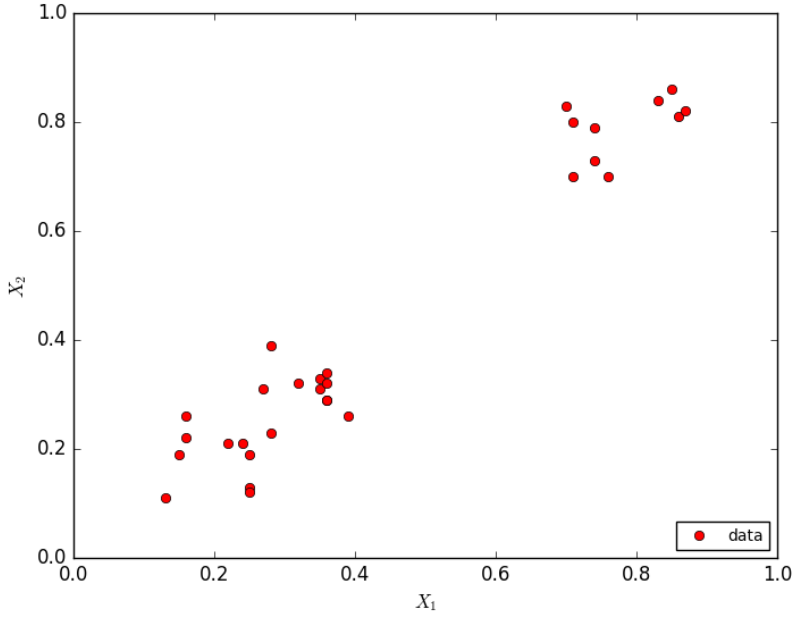

Figure 2: Example clustering dataset

4. **Theory (15 pt)**

   (a) **(4 pt)** Imagine we want to apply a perceptron and a multi-layer neural network to a dataset. Which of these two approaches would have a higher VC dimension? Argue why.

   (b) **(3 pt)** What does the VC dimension tell you about PAC learnability?

   (c) **(4 pt)** Which one of the two (perceptron and multi-layer neural network) would need more data to generate a low (we are minimizing the error) out-of-sample error? Argue why.

   (d) **(4 pt)** If we would have an infinite amount of data, which learning approach (i.e. perceptron or multi-layer neural network) would you expect to work best? Argue why.

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

   Imagine the following dynamical systems model of the relationship between stamina (how much endurance a person has) and the intensity of activities being conducted:

$$\hat{y}_{stamina}(t + \Delta t) = y_{stamina}(t) + \gamma \cdot (y_{activity\_level}(t) - y_{stamina}(t)) \cdot \Delta t \quad (1)$$

$$\hat{y}_{activity\_level}(t + \Delta t) = y_{activity\_level}(t) + \gamma \cdot \Delta t \quad (2)$$

   The model basically says that *stamina* increases when the *activity level* is above the current *stamina*. The precise change depends on the parameter γ. The *stamina* decreases when

the *activity level* is below the current *stamina*. Furthermore, the *activity level* increases with a fixed value γ. We assume a setting of $\Delta t = 1$. In addition, we have collected a dataset shown in Table 1 about the values for *stamina* and the *activity level*. Finally, we assume the absolute difference to be used as a distance metric (i.e. $E(target) = \sum_{t=0}^{N} |\hat{y}_{target}(t) - y_{target}(t)|$).

Table 1: Example dataset

Time point	*Stamina*	*Activity level*
0	0.5	0.1
1	0.4	0.2
2	0.3	0.3

(a) **(3 pt)** List three machine learning algorithms that have been treated during the lecture and can be used to optimize the parameter λ of the dynamical systems model on the data.

(b) **(5 pt)** Assume we want to predict both *stamina* and *activity level* well. Give an example of two parameter settings for the dynamical systems model whereby one model instance clearly dominates the other model instance when considering the data in Table 1. Explain your reasoning.

(c) **(7 pt)** Give an example of three model instances (i.e three different parameter settings) that do not dominate each other but do have different scores in terms of the two targets. Explain your reasoning.

(d) **(5 pt)** If we were to use a time series based approach, such as ARIMA, instead of the dynamical systems model, what parameters would we need to optimize to accurately describe the data? And what do these parameters represent?