# Exam Machine Learning for the Quantified Self

28. 06. 2019
12:00 - 14:45

NOTES:

1. YOUR NAME MUST BE WRITTEN ON EACH SHEET IN CAPITALS.
2. Answer the questions in Dutch or English (English is preferred).
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 (20 pt)**

   Sylvester is a social guy (despite his rather limited skill to articulate words properly) who likes to challenge his friends. Being quite wealthy he tends to buy the latest gadgets that can make any contribution to understanding and working on his physical shape such as smart watches, chest streps that measure respiration and heart rate, etc.. He then shares the results of such measurements in his network of friends and challenges them to see who is in best shape.

   (a) **(5 pt)** Gimpel *et al.* identify the five factor framework of self tracking motivations. List the five factors and argue which purpose best fits Sylvester.

   (b) **(3 pt)** To gain insight into the differences between Sylvester and his friends we are going to apply clustering. Explain what learning setup for clustering would be appropriate (in terms of data) to identify different groups among the network of friends of Sylvester.

   While Sylvester is still a healthy guy, he suffers from an addiction to substances that are not good for his health. His doctor recommended Sylvester to use an app which assists him in battling the addiction. This app will take the measurements from the

smart phone and all Sylvester's gadgets and will send interventions when appropriate. Assume the app uses a Reinforcement Learning algorithm to determine when to send an interventions.

(c) **(3 pt)** Specify in words what a reward function could look like for the specific case described above.

(d) **(3 pt)** Define a possible state space that would be relevant for the task at hand. Argue why the state space is appropriate.

(e) **(3 pt)** Let us assume that the state space that we have defined is continuous. Would we be able to apply the off-the-shelf Q-learning approach we have discussed during the lecture? Explain why (not).

(f) **(3 pt)** Explain the difference between on policy and off policy Reinforcement Learning.

2. **Outlier Detection (20 pt)**

(a) (**3 pt**) Explain the difference between distance based outlier detection and distribution based outlier detection.

Consider the data shown in Figure 1.

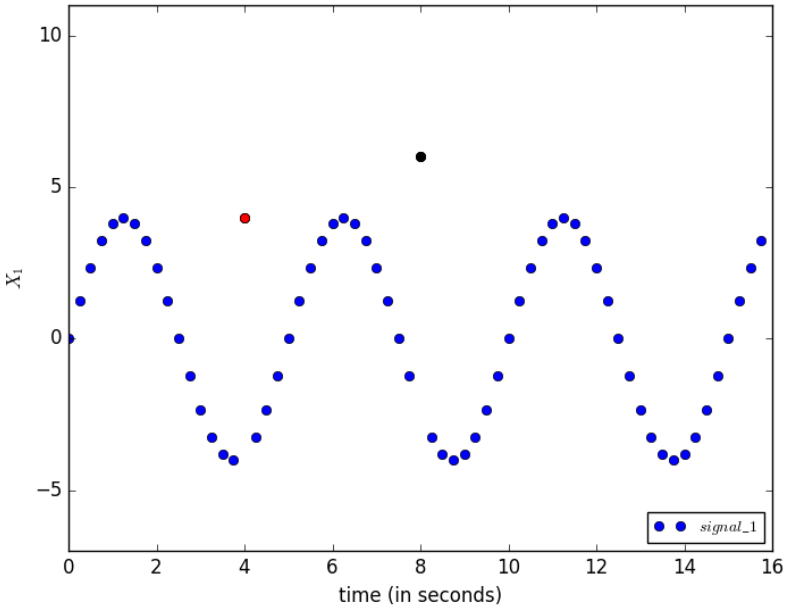


Figure 1: Example dataset - outlier

(b) (**5 pt**) Let us assume we use Chauvenet's criterion to detect outliers with respect to attribute $X_1$. When we focus on the red measurement in Figure 1 (at 4 seconds, with value 4), do you think this point would be considered an outlier according to this algorithm? And the black measurement at 8 seconds with value 6? Argue for both cases why (not).

(c) (**4 pt**) We will apply a very simple version of the Kalman filter (with just identity matrices) to this data to detect outliers and impute the data. Explain what the dataset after the application of the Kalman filter to detect outliers and impute values would look like.

(d) (**4 pt**) We see that the measurements in this dataset are relatively coarse grained. We want to impute their values using either a mean value imputation or using interpolation. Argue which approach would be most suitable for the case at hand.

(e) (**4 pt**) Someone argues that we should apply a lowpass filter to get rid of potential noise in our data. This person suggests to use a cut-off frequency of $f_c = 0.1Hz$ with a very high order of the filter (i.e. value for $n$). What would the result of the application of the filter be? Describe your result in words or draw a graph. Explain how you came to your answer.

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

Consider the data shown in Figure 2.

(a) **(5 pt)** We apply a Fourier transformation to both signals ($signal\_1$ and $signal\_2$) and use the frequency with the highest amplitude as a feature. Which of the two series would have the highest value for this feature? Argue your choice by explaining how you infer the value of this feature for both series.

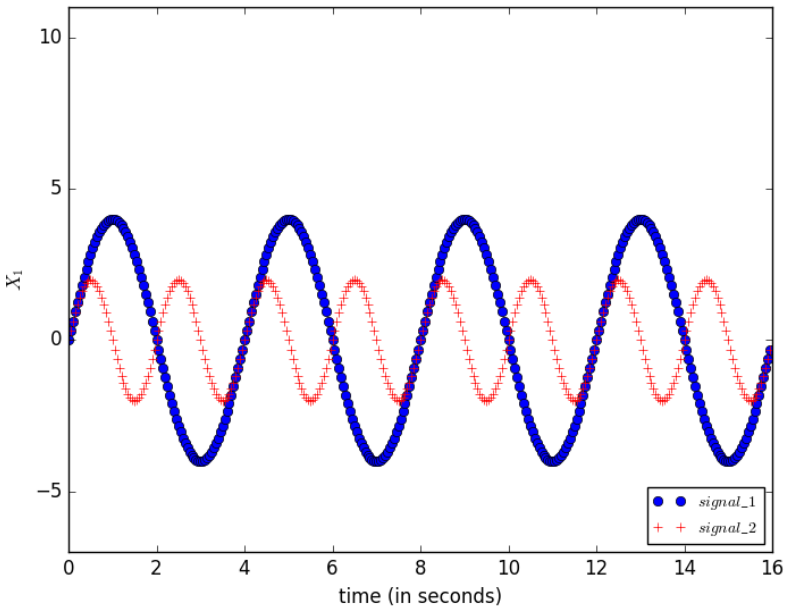


Figure 2: Example dataset - temporal

(b) **(4 pt)** From now on, assume we set our step size for the dataset to be $\Delta t = 1$ *second*. We apply an approach from the time domain, namely we take the mean value over a window size of $\lambda = 3$. Compute what the value of the feature in the time domain would be for the different time points we have available and do so for both signals. Explain how you came to your results.

(c) **(4 pt)** In feature engineering we can also consider other types of data next to the numerical data we have discussed so far. Explain the difference between a bag-of-words approach to engineer features from text compared to a topic modeling approach.

Table 1: Example dataset - temporal with categorical data

| *Time point* | *Mood* | *Activity level* | *Depressed* |
|---|---|---|---|
| 0 | positive | high | no |
| 1 | positive | high | no |
| 2 | positive | low | no |
| 3 | positive | high | yes |
| 4 | negative | high | yes |

(d) (**7 pt**) Apply the algorithm as proposed by Batal *et al.* to extract temporal features in the time domain on the combination of the categorical features *Mood* and *Activity level* on the data shown in Table 1. Consider a window size $\lambda = 1$ and a support threshold of $\Theta = 4/4$ (i.e., 1, this is the minimum support needed). Explain what features result. Explain how you came to these features.

4. **Clustering (15 pt)**

In this question, we are going to focus on comparing dataset of different quantified selves using clustering.

(a) (**4 pt**) Name the two conditions we have for matching data points in dynamic time warping and explain what they entail.

(b) (**3 pt**) When we use dynamic time warping as a distance metric between different datasets, what clustering algorithm is preferred: k-mean clustering or k-medoids clustering? Explain why.

(c) (**4 pt**) When we have a large number of features, clustering might not provide us with very insightful results. Explain how the subspace clustering algorithm tries to mitigate this problem.

(d) (**4 pt**) Name and explain one metric that can be used to evaluate the quality of clustering that has been treated during the course.

5. **Supervised Learning (15 pt)**

Assume we have a time series available covering four different measurements (attributes) and one target we want to predict, which is numerical. We have 100 time points of data available where we both have values of the attributes and have target information available.

(a) **(4 pt)** In trying to learn based on this problem, one person suggests a recurrent neural network with four hidden layers, each including 1000 neurons, to make sure

the model can potentially capture all interesting relationships. Explain by means of the concepts from the learning theory we have discussed (e.g PAC learnability) whether this would be a suitable choice given the described dataset.

(b) **(3 pt)** Explain the difference in training procedure between a regular recurrent neural network and an echo state network.

(c) **(2 pt)** We are going to apply an echo state network to this problem. What would be the range you would try for setting the number of neurons inside the reservoir?

(d) **(3 pt)** We have a dynamical systems model that predicts the next vale of our target based on the current measurement and our attributes. The model has several parameters we can set, which influence the way in which the predictions are made. Give three approaches that can be used to select the best values of these parameters and have been discussed during the course.

(e) **(3 pt)** We are going to learn on this dataset (of one individual) using an algorithm, how can we best split our dataset into a training, validation and test set? Explain

why.