Resit Machine Learning for the Quantified Self
09. 08. 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)**

   Steve is a student with serious health problems. He is obese due to a lack of movement and regularly has mental health problems (depression, anxiety). He is pretty addicted to his mobile phone which is equipped with all imaginable sensors.

   (a) **(4 pt)** Provide the definition of the quantified self that has been discussed during the lecture.

   (b) **(5 pt)** Identify a supervised machine learning task and an unsupervised machine learning task that could be useful for the case of Steve.

   (c) **(3 pt)** List three measurements that could be collected for the case of Steve and argue their relevance based on one or both tasks you have identified above.

   (d) **(3 pt)** Provide the definition for reinforcement learning that has been treated during the lecture.

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

   Consider the data shown in Table 1.

   (a) **(4 pt)** Apply a transformation in the time domain for the attribute *Activity level* using a mean aggregation function. Use a window size $\lambda = 1$. Provide the values for the newly created attribute.

   (b) **(6 pt)** Apply the algorithm proposed by Batal *et al.* with a window size $\lambda = 1$ and a minimum support of $\theta = \frac{2}{7}$ on the attribute *Activity type*. List what new attributes result and show your calculations.

   (c) **(4 pt)** Next to the time domain, which other domain is available for aggregation of temporal features? Discuss how features are created in that domain.

Table 1: Example dataset

Time point	*Activity level (0-100)*	*Activity type*	*Speed*	*Tired*
0	5	sitting	0	no
1	80	running	10	no
2	80	running	9	yes
3	50	walking	5	no
4	80	running	9	no
5	50	walking	5	no
6	80	running	9	no
7	80	running	8	yes

(d) **(6 pt)** Sometimes we have unstructured data. Describe the pipeline you can use to identify features from natural language.

3. **Clustering (20 pt)**

We have collected activity data for two quantified selves, see Table 2. We are going to apply clustering to this data.

Table 2: Two datasets

Time point	*Activity data*
	person 1
1	60
2	60
3	80
4	80
5	60
	person 2
1	60
2	80
3	80
4	60
5	60

(a) (**3 pt**) We are going to cluster on a person level using this data. Explain what a feature-based distance metric is on this person level.

(b) (**5 pt**) Compute the distance using the Euclidean distance function on a person level (an example of a so-called raw-based person level distance metric). Show your calculations.

(c) (**8 pt**) Let us now use dynamic time warping as a distance metric. Fill in Table 3 by using the dynamic time warping algorithm. Use the absolute difference between the values as distance metric. Show the steps you used in the calculations.

Table 3: answer table

person 2	t=1					
	t=2					
	t=3					
	t=4					
	t=5					
		t=1	t=2	t=3	t=4	t=5
		person 1				

(d) **(4 pt)** Given the characteristics and features of the simple dataset above: would subspace clustering be appropriate to use with this dataset? Argue why (not).

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

This question concerns the temporal machine learning algorithms that have been discussed during the lectures.

(a) **(4 pt)** Explain what stationarity means in the domain of time series.

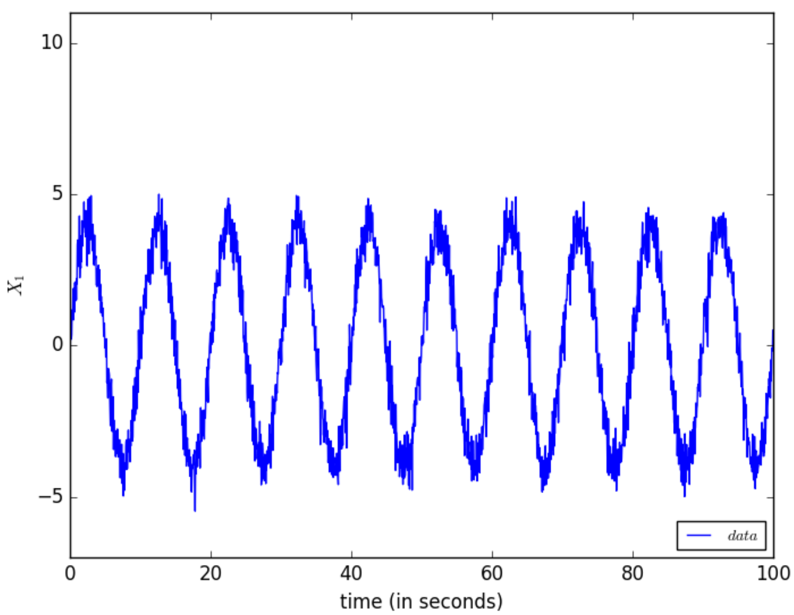

Figure 1: Example dataset

(b) **(4 pt)** Consider Figure 1, which shows a series of a measurement we are trying to predict. Which one of the three components that we normally break a time series down in would explain most of the pattern we see? Argue your choice.

(c) **(5 pt)** We have collected data about three attributes, namely the x, y, and-z-axis of the accelerometer and aim to predict the current activity. We plan to use an echo state network for this purpose. Make a design for an echo state network for this problem and indicate the dimensions of the weight matrices.

(d) **(4 pt)** As an alternative to the echo state network, we also plan to use a regular recurrent neural network. Provide two differences between echo state networks and recurrent neural networks.

(e) **(3 pt)** Regular recurrent neural networks are known to suffer from several problems when learning time series. Provide one such problem and explain what causes this problem.

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

We are going to apply reinforcement learning to support a user in becoming more active. We measure the activity level and activity type of a person and want to provide suggestions to that person based on his measured state (examples of advices could be: do activity x, stop activity y, etc.).

(a) **(3 pt)** Explain what the Markov Property means (you can relate your explanation to this specific example or you can also explain it in general if you want).

(b) **(4 pt)** Explain how the one step Q-learning algorithm works.

(c) **(4 pt)** Some of the measurements we perform are continuous (specifically, the activity level is), would this be a problem for SARSA or Q-learning? Argue why (not).

(d) **(4 pt)** We have the choice to either apply an ϵ-greedy approach or a softmax approach to select the actions. We know that the person we are supporting does not change at all in terms of responses to messages. Which one of the two approaches would be most suitable to use? Argue your choice.