# Exam Machine Learning for the Quantified Self
19. 07. 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 (15 pt)**

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

   Let us consider a specific application of the quantified self. We want to improve the adherence of medication intake. In order to do so, we build an app that learns how to coach a user in such a way that in the long run adherence is high and stable.

   (b) **(4 pt)** When considering the different types of learning problems in machine learning, what type would you consider the learning problem described above? Argue why.

   (c) **(4 pt)** In the development of the aforementioned app, we collect a lot of data from different sensors. These are collected at very different sampling rates. Explain in a step-by-step way how we can create a suitable dataset from such data in order to apply machine learning.

(d) **(3 pt)** Explain two advantages and one disadvantage of considering measurements at a lower level of granularity compared to a high granularity.

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

(a) (**3 pt**) Provide the definition of an outlier as was discussed during the course.

(b) (**5 pt**) Explain the Local outlier factor algorithm on a conceptual level.

Let us consider one sensor $X_1$ for which we measure the following values:
$[1, 1, 2, 2, 3, \mathbf{10}, 20, 20, 21, 22, 22, 23, 34]$
Let us focus on the measured value **10**.

(c) (**9 pt**) For each the following three algorithms argue whether the value **10** would be considered an outlier given the measured values: *Chauvenet's criterion*; A *mixture model* with $K = 2$, and a *simple distance based* approach with $d_{min} = 5$ and $f_{min} = 0.9$. Provide an argumentation for each of the three answers you gave.

(d) (**3 pt**) We suffer from high frequency periodic noise in our dataset. What technique could we use to remove this noise? Argue you choice.

3. **Clustering (20 pt)**

We have collected heart rate data for two quantified selves namely Ali and Steven, see Table 1. We are going to apply clustering to this data.

Table 1: Two datasets

| *Time point* | *Heart rate* |
|---|---|
| | *Ali* |
| 1 | 80 |
| 2 | 80 |
| 3 | 100 |
| 4 | 80 |
| 5 | 80 |
| | *Steven* |
| 1 | 80 |
| 2 | 100 |
| 3 | 80 |
| 4 | 80 |
| 5 | 100 |

(a) (**4 pt**) Imagine we use a feature based approach to compute the distance between the time series using the mean. Compute the distance between the two time series using the Euclidean distance.

(b) (**4 pt**) Next to the Euclidean distance, we can also use Gower's similarity as a distance metric. Explain how Gower's similarity is defined.

(c) (**8 pt**) As an alternative, we can apply Dynamic Time Warping (DTW). Fill in Table 2 (next page) by using the Dynamic Time Warping algorithm. Use the absolute difference between the values as distance between two points. Show the steps you used in the calculations.

Table 2: answer table

| | | | | | | |
|---|---|---|---|---|---|---|
| *Steven* | t=5 | | | | | |
| | t=4 | | | | | |
| | t=3 | | | | | |
| | t=2 | | | | | |
| | t=1 | | | | | |
| | | t=1 | t=2 | t=3 | t=4 | t=5 |
| | | *Ali* | | | | |

(d) (**4 pt**) Explain the difference between agglomerative and divisive clustering. Furthermore, explain in this context what a dendrogram is.

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

In this question, we are going to focus on supervised learning based on Quantified Self data. We want to apply a supervised learning approach on sensory data (specifically accelerometer data) to recognize activities. We consider both algorithms that take the notion of time into account and those that do not.

(a) (**4 pt**) Explain for which of the two (machine learning algorithms that take time into account explicitly or not) the temporal feature engineering step is most important.

(b) (**4 pt**) Name the two domains we can identify temporal features in, and explain the difference between the two domains.

(c) (**2 pt**) When considering temporal features using a window combined with non-temporal machine learning approaches, we mostly remove instances of which the windows overlap too much (e.g. we allow for a maximum of 50% overlapping windows). Give one advantage and one disadvantage of allowing more overlap between windows.

(d) (**3 pt**) We need to decide on the algorithm to use. Explain based on the supervised learning theory how the size of the dataset and the complexity of the hypothesis space resulting from the choice of our machine learning algorithm are related.

(e) (**4 pt**) List two approaches that can be used to avoid overfitting of the data. For both approaches, explain how these approaches contribute to less overfitted models.

(f) (**3 pt**) What is the main disadvantage of regular recurrent neural networks when it comes to learning from temporal data? What is the cause of this problem?

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

We are going to focus on a reinforcement learning case. We focus on a case where we want to support people that are hopelessly out of shape to regain their shape again by sending them messages to coach them in their work out endeavors. The MDP which includes the states, actions, and rewards is shown in Figure 1. We see two actions, namely an advice to *work out* and an advise to *remember: start slow*.

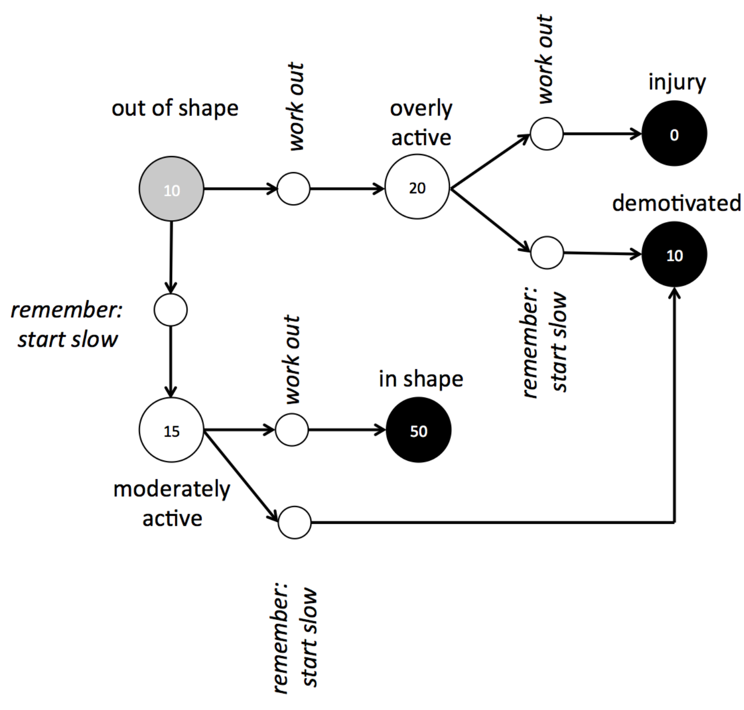


Figure 1: MDP example

(a) **(5 pt)** Assume that we use $\gamma = 1$ to compute the value function. Compute the final Q-values for each state-action pair for the non-terminal states (i.e. you do not have to compute the values for the black nodes). Explain how you came to your answer.

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

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

(d) **(3 pt)** What is the purpose of the U-tree algorithm? Explain how the algorithm works.