Exam Machine Learning for the Quantified Self

29. 06. 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 (15 pt)**

   Cristiano is a sporty guy, and this is also his main occupation. In order to prepare well for an important tournament he is engaged in the quantified self. He tracks a variety of things, including all kinds of physiological measurements (such as heart rate and respiration), his movements (using the sensors of his mobile phone) as well as the training sessions he performs.

   (a) **(3 pt)** Choe et al distinguishes various purposes why someone would engage in the quantified self. Argue which purpose best fits Cristiano.

   (b) **(4 pt)** Identify a supervised machine learning task and an unsupervised machine learning task that could be useful for the case of Cristiano.

   (c) **(5 pt)** Explain for one of the two tasks you have identified above what the table **X** would look like (explain both the columns and the rows).

(d) **(3 pt)** Explain how we could apply reinforcement learning to the case of Cristiano.

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

Consider the data shown in Figure 1.

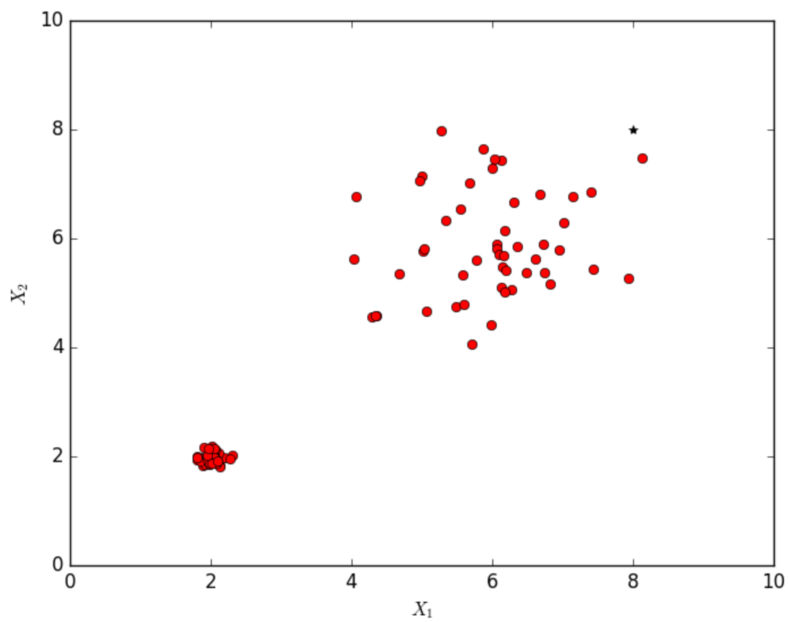

Figure 1: Example dataset

(a) **(4 pt)** We want to apply an outlier detection algorithm to this data. If we consider removing the outliers for the attribute X_1 alone, would you prefer to use Chauvenet's criterion or a mixture model? Argue your choice.

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

(c) **(4 pt)** Let us consider the point shown by means of the black star (at $X_1 = 8$ and $X_2 = 8$). Would it be more likely that this point would flagged as an outlier using the simple distance based outlier detection or using the local outlier factor? Argue why.

(d) **(3 pt)** In outlier detection, the outlier detection algorithms have parameters to be set that eventually influence what is considered to be an outlier or not. Explain how appropriate parameter values can be found.

(e) **(4 pt)** We want to apply a Principal Component Analysis to this data. Illustrate graphically what the principal component would look like. Argue why you have drawn it in that way.

3. **Feature Enginering (15 pt)**

Consider the data shown in Table 1.

Table 1: Example dataset

Time point	*Heart rate*	*Intensity*	*Activity label*	*Tired*
0	60	low	sitting	no
1	60	low	sitting	no
2	70	low	walking	no
3	90	high	walking	yes
4	90	high	walking	yes

(a) (**3 pt**) Given this dataset, we are considering to aggregate the values of heart rate in the time domain by using the mean. A proposal is done to apply a window size of $\lambda = 3$. Given the size of the dataset shown, do you think this an appropriate choice? Argue why (not).

(b) (**8 pt**) Apply the algorithm as proposed by Batal *et al.* to extract temporal features in the time domain on the combination of the features *Intensity* and *Activity label.* Consider a window size $\lambda = 1$ and a support threshold of $\Theta = 3/4$ (this is the minimum support needed). Explain what features result. Explain how you came to these features.

(c) (**4 pt**) Explain how we can derive temporal features in the time domain in case we have a combination of numerical and categorical features.

4. **Clustering (20 pt)**

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

Table 2: Two datasets

Time point	*Heart rate*
	person 1
1	60
2	60
3	60
4	60
5	60
	person 2
1	100
2	120
3	80
4	60
5	60

(a) (**4 pt**) Imagine we use a raw-based approach to determine the distance between the two time series. First, let us consider the Euclidean distance by paring up each data point (i.e. we use time point 1 from person 1 and time point 1 from person 2, etc.). Calculate the distance between the two time series.

(b) (**4 pt**) As an alternative we can apply Dynamic Time Warping (DTW). Explain what the boundary and monotonicity condition are in DTW.

(c) (**8 pt**) Let us now apply DTW. Fill in Table 3 (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 3: answer table

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

(d) (**4 pt**) Explain the difference between k-means and k-medoids clustering. Which of these two algorithms would be best to combine with DTW as distance metric? Explain why.

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

This question concerns the several supervised learning algorithms as well as the theory underlying supervised learning. We assume a learning problem where sensory data ($p = 100$ features) has been collected for $N = 1000$ time points (i.e. learning instances) and we want to create a predictive model for the activity type (similar to our CrowdSignals data).

(a) **(3 pt)** We decide to apply a convolutional neural network with in total 2000 hidden neurons. We nicely split the data into a training and test set. Even though we vary the parameter settings of the network a lot (though not the number of neurons) we experience a lot of overfitting. Explain why this is not surprising. Support your answer by means of the theories that have been treated during the course.

(b) **(5 pt)** We want to select a number of features before we apply a machine learning problem to come to a model. Explain how forward selection works to select features.

(c) **(3 pt)** Explain the concept of regularization and explain how it can be used to avoid overfitting.

(d) **(3 pt)** We are considering the application of one of the temporal learning algorithms that have been discussed in the book to the case at hand (i.e. predict the activity type, a categorical feature). Based on your knowledge of the algorithms, would it be better to apply a *time series algorithm* or a *recurrent neural network*? Argue why.

(e) **(6 pt)** Another way to predict based on temporal patterns involves dynamical systems models. Provide three algorithms that have been treated during the course to learn the best parameter values for such models and briefly explain how they work.