Exercise 2: Rule-based and data-driven examples

Data Description:
This exercise uses one patient from the Improve dataset [1]. This dataset has been collected in an intensive care unit (ICU) at a Finnish university hospital during 1995*. This dataset contains a 24-hour recording of each patient, their vital signs recorded from a patient monitor, computed values calculated several times a day, and manual annotations from the study: a study nurse noted all the activities and events occurring at the patient's bedside. The annotations also include diagnoses and findings given by the ICU doctor.

In this exercise we examine one of the study patients more closely. In more detail, we will examine the signs of cardiac failure for this patient. Cardiac failure is a condition where the heart does not pump enough blood for the body's needs. It can either happen when the heart does not pump enough blood, or the heart is too weak to pump properly. Heart failure does not mean that the heart stops pumping completely (that is called asystole).

Cardiac failure is determined separately for both sides of the heart. There are three different forms / diagnostic criteria for cardiac failure:

1. Inadequate flow and metabolic signs of tissue hypoxia
2. Inadequate flow, no signs of tissue hypoxia
3. Acceptable flow and continuous need of exceptional support

In this case, we determine cardiac failure based on three measured variables: cardiac index, pulmonary capillary wedge pressure (PCWP), and peripheral temperature. All these have been automatically measured or calculated from the ICU data in (almost) real time.

The data for this patient has been preprocessed from the original binary *.edf format to a more easy to read *.csv format. Patient information is in Appendix A

More information about the study:

[1] M. Van Gils, H. Jansen, K. Nieminen, R. Summers and P. R. Weller, "Using artificial neural networks for classifying ICU patient states," in IEEE Engineering in Medicine and Biology Magazine, vol. 16, no. 6, pp. 41-47, Nov.-Dec. 1997, doi: 10.1109/51.637116. available: https://ieeexplore.ieee.org/document/637116

* note that even though the measurements may seem 'ancient', from 1995, the format, accuracy, and other specifications of the data are the same as what you would get if you would do a similar measurement today. Clinically deployed systems continue their lifespan over a long time, and newer generations of devices are backwards compatible.

Tasks

1. Determine a rule-based model for cardiac failure. From the data collected in this study, there are three measured/computed variables located in AN1 file that may be examined:
    a. Cardiac Index (CI), [l/min/m2], variable id = 30001000
    b. Peripheral Temperature (Tp), [degr.C], variable id = 400
    c. Pulmonary capillary wedge pressure (PCWP), [mmHg], variable id = 800

   Note: there are some other indications of cardiac failure, which cannot or have not been measured at this study. That's why this rule-based model is somewhat simplified

   Question: Does the number of annotations for different computed or measured patient health parameters differ? If it is different, explain, why it is so.

   Suggested implementation:
   - use *pandas* for data manipulation (https://pandas.pydata.org/docs/user_guide/index.html)
   - use *DataFrame.loc* and *DataFrame.shape* properties (https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.loc.html and https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.shape.html)

2. Create a rule-based code that checks whether any of the rules below apply to the data, and create a vector containing *1* if cardiac failure is present, or *0* if it is not.
   - CI < 2.0, or
   - Tp < 32.5, or
   - PCWP > 10

   Make a time plot that indicates presence of cardiac failure detected by this rule-based model. Plot should have time in HH:MM in the x axis, which allows you to see the time correctly.
   Question: How reliable is the model based cardiac failure detection? Motivate your answer.

   Suggested implementation:
   - use *Matplotlib* for plotting (https://matplotlib.org/stable/tutorials/introductory/pyplot.html)
   - use *pandas datetime* (https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.to_datetime.html)

3. Define a new vector with cardiac failure annotations from manual annotations (AN2 file) where the 'Variable' indicates a cardiac failure (see Appendix B) and its value denotes the continuation of exceptional support.
   Plot these filtered annotations together with the rule-based diagnosis into a plot of your choice. The aim is to make an informative plot(s) or other type of comparison, where you can compare the results. All plots should have time in HH:MM in the x axis, which allows you to see the time correctly.

   Question: Compare the results of rule-based model and manual annotations, are they same or different? Think about the reasons why these results might differ from each other.

4. Check if annotations for different care activities (Annotations in appendix C) correlate with cardiac failure related annotations.

   Question: How could a time dependent comparison of care activities with a time dependent diagnosis be used? May it help to improve the rule-based models? Motivate your answer.

Appendices

Appendix A: Patient background

The patient examined in this exercise:

- Age: 59
- Gender: Male
- ICU outcome; survived after 32 ICU days
- ICU days before study: 4
- Weight: 61 kg
- Height: 171 cm
- BSA: 1.71 m2
- Admission diagnosis: postoperative ventilatory failure because of respiratory infection, atrial fibrillation (post-operative)
- ICU diagnoses: Acute ventilatory failure, Status post AMI-B (Acute myocardial infarction), Cholecystitis acuta, Pneumonia bact.alia, duodenitis nud, Cholecystectomia (13 of oct., 1995), tracheostomia (16 of oct., 1995), atrial flutter/fibrillation, bradycardia and total block and asystolia.

Data:

Filename	**Contents**
annotation1 .AN1	The file contains annotations for automatic variables. It starts with 2 rows of header: • (i) patient metadata, date and time of the start of observation • (ii) header for the following table The tabular data starts form the row 3. The columns contain the following information: • (i) date time stamp (8 digit number, every 2 digits for month, day, hour, minute respectively) • (ii) variable type ▪ MV monitored variable ▪ CV computed variable (derived) ▪ DV drug volume (quantity) ▪ DR drug infusion rate ▪ FR fluid infusion rate • (iii) variable • (iv) variable's value • (v) status
annotation2 .AN2	The file contains annotations for manual variables. It starts with 2 rows of header: • (i) patient metadata, date and time of the start of observation • (ii) header for the following table The tabular data starts form the row 3. The columns contain the following information: • (i) date time stamp (8 digit number, every 2 digits for month, day, hour, minute respectively) • (ii) variable type ▪ CA care activity ▪ LB lab variable ▪ OV observed variable • (iii) variable • (iv) variable's value • (v) status
info.txt	Information file of data channels.

	Format of file: ['Fs','Start_date','Start_time','Label','Dimension','Coef1','Coef2','Coef3','Coef4','Nmb_chans','N']

Appendix B: Annotation codes

AN2 codes to be found:

- 15001114: Left cardiac failure: inadequate flow + metabolic signs of tissue hypoxia
- 15001115: Left cardiac failure: inadequate flow + no signs of tissue hypoxia
- 15001116: Left cardiac failure: acceptable flow + cont. need of exceptional support
- 15001117: Right cardiac failure: inadequate flow + metabolic signs of tissue hypoxia
- 15001118: Right cardiac failure: inadequate flow + no signs of tissue hypoxia
- 15001119: Left cardiac failure: acceptable flow + cont. need of exceptional support

Example of the values:

- 15001116,1,Left cardiac failure: acceptable flow + cont. need of exceptional support, No
- 15001116,2,Left cardiac failure: acceptable flow + cont. need of exceptional support, Yes
- 15001116,3,Left cardiac failure: acceptable flow + cont. need of exceptional support, Can't be assessed
- 15001116,4,Left cardiac failure: acceptable flow + cont. need of exceptional support, ...

The annotated disorders related to annotation of cardiac failure:

Disorders	**Start**	**End**
Left cardiac failure IIIC	10:00	05:03
Cardiac index < 2.0	11:54	12:42
Tp < 32 (degrees Celsius)	12:37	14:36
	19:41	01:14

Appendix C: Annotation of care activities

- 82,-1,Low CI and filling norm .
- 83,-1,Low CI and filling high or normal (rcf)
- 84,-1,Low urine output (lcf) .
- 85,-1,Low urine output (rcf) .
- 86,-1,Low Tp, CI, filling normal or high (cf)
- 87,-1,Low Tp, CI, filling normal or high (rcf)
- 88,-1,CI normal or low, needs high fP (cf)
- 89,-1,CI normal or low, needs high fP (rcf)
- 90,-1,CI normal or low, fP high or normal, vasodil (cf)
- 91,-1,CI normal or low, fP high or normal, vasodil (rcf)
- 92,-1,CI normal or low fP high or normal inos (cf)
- 93,-1,CI normal or low fP high or normal inos (rcf)
- 116,-1,Left cardiac failure
- 117,-1,Right cardiac failure