Quiz M7.02

Note: For each question **make sure you select all of the correct options—** there may be more than one! Don't forget to use the sandbox notebook if you need.

Question 1 (1/1 point)

We have a dataset with patient records from 10 different hospitals, and our goal is to predict whether a patient has a disease or not. Let's also suppose that the classes ("disease" and "no-disease") are imbalanced. Additionally, we suspect that each hospital's data may have systematic biases due to factors like medical devices, policies, socioeconomic status of the patients, etc.

Which cross-validation strategy is the most suitable for assessing the model's ability to make good predictions on patients from hospitals not seen during training?

- (●) a) Group stratified k-fold cross-validation ✔
- ( ) b) Group k-fold
- ( ) c) Stratified k-fold cross-validation
- ( ) d) Leave-one-out cross-validation

EXPLANATION

solution: a)

Different hospitals represent different groups. It could be the case that different hospitals have significant biases in the population of their patients as mentioned above. Evaluating a machine learning model without taking this information into account would lead to over-optimistic results.

similar to that of the original dataset). Thus, a group-aware, class-stratified strategy takes both the grouping and class imbalance of the dataset into account.

You have used 2 of 2 submissions

YOUR EXPERIENCE

According to you, this whole 'Choice of cross-validation' lesson was:

- ○ **Too easy, I got bored**
- ○ **Adapted to my skills**
- ○ **Difficult but I was able to follow**
- ○ **Too difficult**

Submit

To follow this lesson, I spent:

- ○ **less than 30 minutes**
- ○ **30 min to 1 hour**
- ○ **1 to 2 hours**
- ○ **2 to 4 hours**
- ○ **more than 4 hours**
- ○ **I don't know**

Submit

FORUM (EXTERNAL RESOURCE)

There are no more M7. Quiz M7.02 topics. Ready to start a new conversation?