Ricardo Aler Mur

SELF-ASSESSMENT OF LEARNING TECHNIQUES WITH QUESTIONS AND ANSWERS

1) What is the model evaluation method most commonly used when little data is available?

Answer. The cross-validation method. If you have few data, the assessment will not be significant and will be very sensitive to random biases that may contain the whole test. Therefore, cross-validation repeats the training cycle-test several times (10 times in the case of 10-fold crossvalidation) and calculates the average at the end, so that random biases get canceled when calculating the average.

2) What are stratified partitions? What kind of problems are they convenient to use?

Answer: They are partitions that are intended to preserve in the partition training and test the proportion of classes observed in the global data set. The aim is that the partitions are representative of the original problem. Often, they are used in problems of imbalanced samples, since in such cases it may happen that by chance, little or no data from the minority class have the appropriate number of representatives in the training partition or test. In extreme cases, if the minority class has very little data, it can happen that training or test partitions do not contain any data of that class if partitions are made randomly.

3) Say an algorithm that allows learning models taking into account the distribution of classes (for imbalanced problems)

Answer: SMOTE: Synthetic Minority Over-sampling Technique. This technique allows resampling the minority class.. It is not merely a replication of the existing instances but it creates new instances in a "reasonable" way and avoids one of the major problems of pure replication, which is overlearning.

4) What is a ROC curve for?

Answer: 1) Used to represent the behavior of a classifier (scorer) for all possible operating conditions. 2) Select the best threshold of a classifier (scorer) for specific operating conditions. 3) Allows comparing, selecting and discarding a set of classifiers for certain operating conditions.