

Business Analytics

Exercise Sheet 3

Nicolas Schilling (schilling@ismll.de)
Information Systems and Machine Learning Lab (ISMLL)
Universität Hildesheim

14 May 2013
Submission Until 26 May 2013 23:59

Question 1: Recall and Precision for Multi-Label Prediction (3 points)

Consider a general multi-label prediction problem. Define Recall and Precision in formulas, and explain the meaning of both quality measures in your own words. For which predictor \hat{Y} is Recall maximized, for which \hat{Y} is Precision maximized?

Question 2: Apply Recall and Precision for Multi-Label Prediction (3 points)

Let us suppose we have a set of movies:

$$\mathcal{M} := \{It, Titanic, The Shining, Terminator, Star Wars: Episode 2, Men in Black\}$$

which we want to classify into three categories, *Action* $\hat{=}$ A , *Horror* $\hat{=}$ B and *Romance* $\hat{=}$ C . Since movies can belong to more than one category, we have a multi-label classification problem. The ground truth for the movies (in the same order as above) is given by:

$$Y := \{y_1 = \{B\}, y_2 = \{C\}, y_3 = \{A, B\}, y_4 = \{A\}, y_5 = \{A, C\}, y_6 = \{A\}\}$$

Let us consider the following prediction:

$$\hat{Y} := \{\hat{y}_4 = \{A\}, \hat{y}_3 = \{B\}, \hat{y}_2 = \{C\}\}$$

Compute Precision and Recall for all three movies where predictions have been made. What is the dependency between Precision and Recall?

Question 3: Boxplots with R (4 points)

In the data set *boxplot.csv*, we have three columns representing 1000 random values of an underlying probability distribution. Feed the data into R and compute all requirements for the boxplot. Then make a boxplot of all three underlying distributions. Which of the three columns represents a standard normal distribution? Give a short explanation, why.

Submission

- Electronically to josif@ismll.de **AND** schilling@ismll.de
- Email title must be *BA2013-NAME-Tutorial-NO*, e.g.: *BA2013- JosifGrabocka-Tutorial-1*
- Report file must be a PDF with file name like the email title *BA2013-NAME-Tutorial-NO.pdf*
- Source codes and other materials must be a ZIP with file name like the email title.