Business Analytics
Exercise Sheet 3

Martin Wistuba (wistuba@ismll.de)
Information Systems and Machine Learning Lab (ISMLL)
Universität Hildesheim

14 May 2013
Submission until 21 May 2013 23:59

Exercise 8: 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?

Exercise 9: Apply Recall and Precision for Multi-Label Prediction (3 points)

Let us suppose we have a set of movies:

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

which we want to classify into three categories, $\textit{Action} \,\hat{=}\, A$, $\textit{Horror} \,\hat{=}\, B$ and $\textit{Romance} \,\hat{=}\, C$. 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?

Exercise 10: 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 wistuba@ismll.de. Text submitted as pdf, code submitted as source files. No archives.