

Business Analytics

Exercise Sheet 2

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

6 May 2013
Submission until 13 May 2013 23:59

Exercise 5: Loss Metrics and Best Predictive Values (3 points)

Enumerate the loss metrics and best corresponding predictive values per loss, for both regression and classification. Fill out the column names and row names (?) of the Tables 1 and 2, where each column name is one type of loss and the respective index row the best predictive value.

| | MSE | ? | ? | ... |
|------|-----|---|---|-----|
| Mean | | | | |
| ? | | | | |
| ? | | | | |
| ... | | | | |

Tabelle 1: Regression - Loss vs Best Predictive Values

| | MCR | ? | ? | ... |
|----------|-----|---|---|-----|
| Majority | | | | |
| ? | | | | |
| ? | | | | |
| ... | | | | |

Tabelle 2: Classification - Loss vs Best Predictive Values

Exercise 6: Apply Loss Metrics and Best Predictive Values (4 point)

Apply the loss metrics and the best predictive values described in Exercise 5 to the data presented in Table 3. Each cell in the table should correspond to the loss value of applying the column-wise metric to the respective row-wise predicted value. For regression you should use the Revenue column data and for classification the Expectations column. Treat Above as 1 and Below as 0.

(Hint: Best values are computed using the training data, while loss metrics are evaluated on test data).

| Year | Revenue (Million \$) | Expectations |
|------------|----------------------|--------------|
| Train Data | | |
| 1998 | 24 | Below Plan |
| 1999 | 32 | Above Plan |
| 2000 | 25 | Below Plan |
| 2001 | 26 | Above Plan |
| 2002 | 24 | Below Plan |
| 2003 | 30 | Above Plan |
| 2004 | 29 | Below Plan |
| 2005 | 39 | Above Plan |
| 2006 | 35 | Below Plan |
| 2007 | 25 | Below Plan |
| Test Data | | |
| 2008 | 28 | Above Plan |
| 2009 | 21 | Below Plan |
| 2010 | 22 | Above Plan |
| 2011 | 28 | Above Plan |
| 2012 | 23 | Below Plan |

Tabelle 3: Regression and Classification Data

Exercise 7: Logistic Loss - Best Predictive Value (3 points)

Logistic Loss, defined in Equation 1, is another popular loss for measuring the prediction accuracy of binary targets. What is the best predictive value for the logistic loss? (Hint: $\frac{\partial L(Y, \hat{Y})}{\partial \hat{Y}} = 0$.)

$$L(Y, \hat{Y}) = \sum_{Y \in D^{Train}} -Y \log(\hat{Y}) - (1 - Y) \log(1 - \hat{Y}) \quad (1)$$

Submission

- Electronically to wistuba@ismll.de. Text submitted as pdf, code submitted as source files. No archives.