Barbora Hladká

Martin Holub

{Hladka | Holub}@ufal.mff.cuni.cz

Charles University,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics
Programming questions

- Student data set
- Most Frequent Classifier
  - $Y = \{1, \ldots, N\}$
  - $D = \{(x_i, y_i), x_i \in X, y_i \in Y, i = 1, \ldots, n\}$
  - most frequent class in $D$: $k^* = \arg\max_{k=1,\ldots,N} \sum_{i=1}^{n} \delta(y_i = k)$
  - $f(x) = k^*$
- Logistic regression glm()
- Accuracy, Precision, Recall, F-measure