

# Introduction to Machine Learning

## NPFL 054

<http://ufal.mff.cuni.cz/course/npfl054>

**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
  - $Y = \{1, \dots, N\}$
  - $D = \{\langle \mathbf{x}_i, y_i \rangle, \mathbf{x}_i \in \mathbf{X}, y_i \in Y, i = 1, \dots, n\}$
  - most frequent class in  $D$ :  $k^* = \operatorname{argmax}_{k=1, \dots, N} \sum_{i=1}^n \delta(y_i = k)$
  - $f(\mathbf{x}) = k^*$
- Logistic regression `glm()`
- Accuracy, Precision, Recall, F-measure