

# Guessing the cuisine of recipes according to their ingredients

Project "pattern recognition" - Pre-presentation

Michael Jobst

October 26, 2015

## Task description

- ▶ guessing the cuisine a recipe belongs to from the list of its ingredients
- ▶ example data: `{'ingredients': ['romaine lettuce', 'black olives', 'grape tomatoes', 'garlic', 'pepper', 'purple onion', 'seasoning', 'garbanzo beans', 'feta cheese crumbles'], 'cuisine': 'greek', 'id': 10259}`

# Task description

- ▶ guessing the cuisine a recipe belongs to from the list of its ingredients
- ▶ example data: {'ingredients': ['romaine lettuce', 'black olives', 'grape tomatoes', 'garlic', 'pepper', 'purple onion', 'seasoning', 'garbanzo beans', 'feta cheese crumbles'], 'cuisine': 'greek', 'id': 10259}
- ▶ task size:
  - ▶ 39774 recipes in the whole dataset
  - ▶ 20 cuisines (→ classes)
  - ▶ 6081 unique ingredients (→ features)
  - ▶ on average 10.8 ingredients per recipe

## Approach for solving

1. convert data (simple import from JSON)
2. feature extraction

## Approach for solving

1. convert data (simple import from JSON)
2. feature extraction
  - 2.1 building sparse feature vectors
  - 2.2 feature selection ( $\chi^2$ -test...)

## Approach for solving

1. convert data (simple import from JSON)
2. feature extraction
  - 2.1 building sparse feature vectors
  - 2.2 feature selection ( $\chi^2$ -test...)
3. selecting and tuning a good classifier

## Approach for solving

1. convert data (simple import from JSON)
2. feature extraction
  - 2.1 building sparse feature vectors
  - 2.2 feature selection ( $\chi^2$ -test...)
3. selecting and tuning a good classifier
  - ▶ SVM
  - ▶ Random tree forests

## Approach for solving

1. convert data (simple import from JSON)
2. feature extraction
  - 2.1 building sparse feature vectors
  - 2.2 feature selection ( $\chi^2$ -test...)
3. selecting and tuning a good classifier
  - ▶ SVM
  - ▶ Random tree forests
4. evaluation



## Approach for solving

1. convert data (simple import from JSON)
2. feature extraction
  - 2.1 building sparse feature vectors
  - 2.2 feature selection ( $\chi^2$ -test...)
3. selecting and tuning a good classifier
  - ▶ SVM
  - ▶ Random tree forests
4. evaluation

### Questions to evaluate

- ▶ Which classifier yields the best results and how to train it efficiently?
- ▶ Is there some interesting statistical information?