ENGSCI 712 – Computational Techniques for Signal Processing

Module I: Feature Extraction, Lecture 6 – Assignment 1.2

Andreas W. Kempa-Liehr

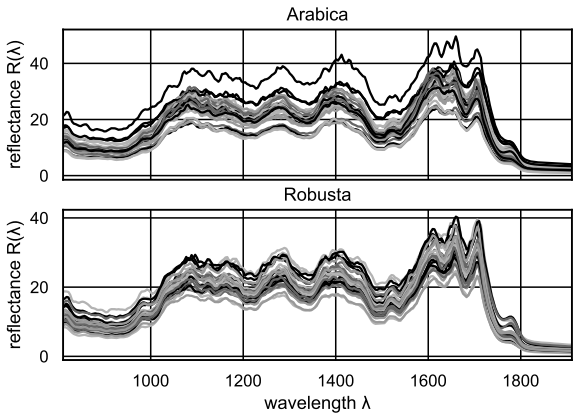
THE UNIVERSITY OF AUCKLAND NEW ZEALAND | ENGINEERING

Department of Engineering Science

05.08.2020 – FE06-CSP.pdf, Rev. 577f3b6

Welcome to CoffeeAI

You have decided to join the Auckland based technology startup CoffeeAI, which develops an *intelligent* coffee machine named *eBarista*. This coffee machine is supposed to automatically tune its grinding and brewing process according to the coffee beans provided.
The heart of each *eBarista* is a Fourier-Transform Infrared Spectrometer, which measures the infrared spectrum of absorption of the coffee beans.
Your task is to use computational signal processing to generates features, which can be used for discriminating Arabica and Robusta beans.



Data source: Romain Briandet, E. Katherine Kemsley, and Reginald H. Wilson. "Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics". In: *Journal of Agriculture and Food Chemistry* 44.1 (1996), pp. 170–174. DOI: 10.1021/jf950305a

In order to evaluate your features, you're going to fit a logistic regression classifier and perform an in-sample analysis.

Fourier Transform Infrared Spectrometer

FTIR spectrometers (Fourier Transform Infrared Spectrometer) are widely used in organic synthesis, polymer science, petrochemical engineering, pharmaceutical industry and food analysis. In addition, since FTIR spectrometers can be hyphenated to chromatography, the mechanism of chemical reactions and the detection of unstable substances can be investigated with such instruments.
Source: `https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Vibrational_Spectroscopy/Infrared_Spectroscopy/How_an_FTIR_Spectrometer_Operates`

FTIR_Spectra_instant_coffee.csv

- Download the `FTIR_Spectra_instant_coffee.csv` from Canvas.
- Load the data set into your Jupyter notebook using the read_csv function from pandas

```
import pandas as pd
data = pd.read_csv(...)
```

- Visualize the data set using function from `seaborn`

```
import seaborn as sns
```

seaborn: statistical data visualization — seaborn 0.10.1 documentation

Design signal processors, which extract information about the statistics of the signals

- Extract the following features from the spectra: minimum, mean, standard deviation, variance, and maximum
- Fit a logistic regression classifier using the L1 norm `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html`
- Inspect the most important features
- Visualize the most important features
- Evaluate the accuracy and confusion of your classifier

Design a signal operator and improve your classifier

- Develop a signal operator, which transforms the original reflectances to new signals,
- Use the output signal to extract the same features as listed above
- Fit a logistic regression classifier using the L1 norm
- Inspect the most important features
- Visualize the most important features
- Evaluate the performance of your classifier

Use a broad range of time series features

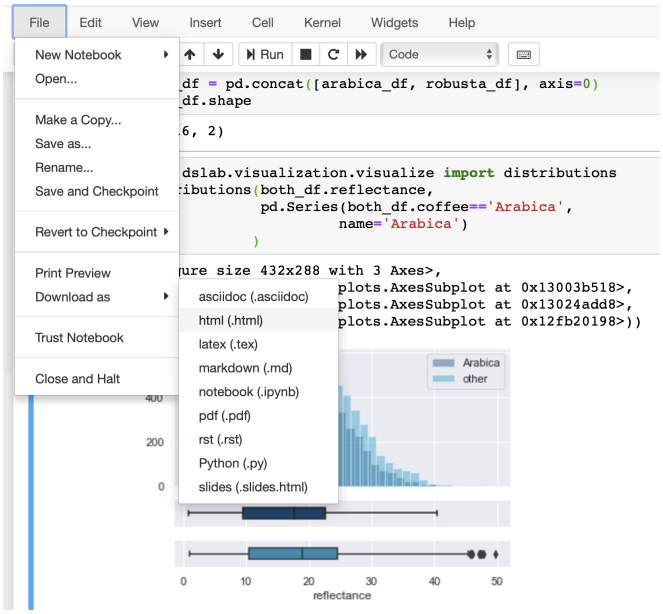
- Use the time series feature engineering library tsfresh for extracting a broad variety of features

```
import tsfresh
```

- Select the statistically significant features
- Retrain the logistic regression classifier
- Evaluate the performance
- Explain and visualize three features of your choice

Write and submit your report

- Export your Jupyter notebook as html



- Upload the html file to Canvas