FEATURE EXTRACTION USING MATLAB IN < 30 MINUTES

PREMANANDA INDIC, PH.D.
DEPARTMENT OF ELECTRICAL ENGINEERING
OUTLINE

- INTRODUCTION
- DIFFERENT TYPES OF FEATURES
- DISCUSSION
OUTLINE

- INTRODUCTION
- DIFFERENT TYPES OF FEATURES
- DISCUSSION
INTRODUCTION

Why do we need to derive features from data?

- To test an hypothesis
- To classify different groups
- To predict outcomes
INTRODUCTION

Source: https://en.wikipedia.org/wiki/30_Minute_Meals
INTRODUCTION

Preprocessing  
Feature Extraction  
Feature Selection

Data

30 Minute Feature Extraction

Statistical or Machine Learning Models

University of Texas at Tyler

INTRODUCTION

Preprocessing
Feature Extraction
Feature Selection

Data

30 Minute Feature Extraction

Text Editor
Workspace
Command window

Statistical or Machine Learning Models
Feature extraction is the process of transforming raw data into numerical features while preserving the information of the original data set.
INTRODUCTION

Feature extraction identifies most discriminating characteristics in signals.
OUTLINE

 INTRODUCTION

 DIFFERENT TYPES OF FEATURES

 DISCUSSION
DIFFERENT TYPES OF FEATURES

- VARIABLE
- STATISTICAL FEATURES
- SPECTRAL FEATURES
- NONLINEAR FEATURES
DIFFERENT TYPES OF FEATURES

- **VARIABLE**

```matlab
lm = fitlm(tbl,'MPG~Weight+Acceleration')
```

<table>
<thead>
<tr>
<th>Weight</th>
<th>Acceleration</th>
<th>MPG</th>
</tr>
</thead>
<tbody>
<tr>
<td>3504</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>3693</td>
<td>11.5</td>
<td>15</td>
</tr>
<tr>
<td>3436</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>3433</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>3449</td>
<td>10.5</td>
<td>17</td>
</tr>
</tbody>
</table>

Linear regression model:  
```
MPG ~ 1 + Weight + Acceleration
```

Estimated Coefficients:
```
                        Estimate       SE         tStat      pValue
(Intercept)              45.155     3.4659    13.028     1.6266e-22
Weight                  -0.0082475  0.00059836 -13.783     5.3165e-24
Acceleration           0.19694      0.14743  1.3359     0.18493
```

\[ MPG = a + b \text{ Weight} + c \text{ Acceleration} \]

Number of observations: 94, Error degrees of freedom: 91
Root Mean Squared Error: 4.12
R-squared: 0.743, Adjusted R-Squared 0.738
DIFFERENT TYPES OF FEATURES

Real Estate Data

```
>> lm=fitlm(housing)

lm =

Linear regression model:
    median_house_value ~ [Linear formula with 9 terms in 8 predictors]

Estimated Coefficients:
                           Estimate     SE      tStat     pValue
__________________________          ____________    ______    __________
(Intercept)         -3.5854e+06     62901     -57.001     0
longitude           -42730        717.09     -59.588     0
latitude            -42510        676.95     -62.796     0
housing_median_age  1157.9         43.389      26.687   2.9463e-154
total_rooms         -8.2497       0.79426    -10.387     3.2948e-25
total_bedrooms      113.82         6.9306      16.423     3.1889e-60
population          -38.386        1.0841     -35.407     1.4597e-266
households          47.701         7.5466      6.3209     2.6535e-10
median_income       40298         337.21      119.5       0
```
DIFFERENT TYPES OF FEATURES

VARIABLE

Don’t want to write the code?
DIFFERENT TYPES OF FEATURES

STATISTICAL FEATURES

- **Mean**: Low Risk and High Risk with N = 128
- **Kurtosis**: Low Risk and High Risk with N = 128
DIFFERENT TYPES OF FEATURES

STATISTICAL FEATURES

Correlation with Self Reported Suicidal Ideation

Mean: \( r = -0.17 \), \( p = 0.05 \)

Variance: \( r = -0.05 \), \( p = 0.53 \)

Skewness: \( r = 0.23 \), \( p = 0.007 \)

Kurtosis: \( r = 0.18 \), \( p = 0.03 \)

MATLAB functions

mean(filename)
variance(filename)
 skewness (filename)
kurtosis(filename)

\[ [r,p]=
corr(resultsfilename(:,1),resultsfilename(:,5)); \]

Atluri, Salvatore and Indic (unpublished)
DIFFERENT TYPES OF FEATURES

SPECTRAL FEATURES

Spectral Features provide frequency-domain metrics on your data. To compute spectral features, you must already have a power spectrum or an order spectrum variable.

**Spectrum**
- Spectrum — Choose from the available spectrum variables. The software brings up the plot of that variable for reference, and converts the plot from log scale to linear scale.

**Spectral Peaks**
- Peak amplitude — Generate a feature based on the amplitude of the peaks.
- Peak frequency — Generate a feature based on the frequency of the peaks.
- Peak value lower threshold — Constrain peak size to exclude low-amplitude peaks. For more information, in Findpeaks, see the MinPeakHeight name-value pair argument.
- Number of peaks — Number of peaks to generate features for. The software selects N most prominent peaks in the chosen frequency band, going in the descending amplitude order. For more information, in Findpeaks, see the NPeaks name-value pair argument.
- Minimum frequency gap — Specify a minimum frequency gap. If the gap between two peaks is less than this specification, the software ignores the smaller peak of the pair. For more information, in Findpeaks, see the MinPeakDistance name-value pair argument.
- Peak excursion tolerance — Specify the minimum prominence of a peak. The prominence of a peak measures how much the peak stands out due to its intrinsic height and its location relative to other peaks. For more information, in Findpeaks, see the MinPeakProminence name-value pair argument.
DIFFERENT TYPES OF FEATURES

- SPECTRAL FEATURES
  - Wavelet transform
  - wavelets(filename)

DIFFERENT TYPES OF FEATURES

SPECTRAL FEATURES

Individuals during major depression phase

$$\begin{align*}
N &= 24 \\
N &= 1
\end{align*}$$

MATLAB functions:

Wavelets
corr

DIFFERENT TYPES OF FEATURES

- SPECTRAL FEATURES

DIFFERENT TYPES OF FEATURES

- Fluctuation Analysis
- Pattern Analysis
- Fractal Analysis
- Information Categorization Approach
- Power Law
- Entropy
- Dimension

DIFFERENT TYPES OF FEATURES

 NONLINEAR FEATURES

Cravings Detection

App Features
- Proprietary method to identify and track stress, cravings, and other biometrics.
- A wearable device vibrates when an event is detected prompting intervention and de-escalation tools via the RAE mobile app.
- Data is pushed in real-time to a recovery team, where they are able to monitor and intervene when help is needed.

DIFFERENT TYPES OF FEATURES

- Additional MATLAB Code

**Physionet:**  https://physionet.org/

**TISEAN:**  https://www.pks.mpg.de/~tisean/
SUMMARY

ORS Research Design & Data Analysis Lab
Office of Research and Scholarship

- Identification of Features
- Develop MATLAB code for feature extraction
- Set up database
- Assist with experiment protocol and data analysis
- Machine Learning

30 Minute Feature Extraction

Data -> Preprocessing -> Feature Extraction -> Feature Selection

Statistical or Machine Learning Models
SUMMARY

The University of Texas at Tyler Center for Health Informatics & Analytics

- Ready to go features

Data → Preprocessing → Feature Extraction → Feature Selection

CHIA FEATURES

Statistical or Machine Learning Models
THANK YOU

Current Students:

Sloke Shrestha, UG
Mohammed Alenazi, Graduate
Pravitha Ramanand, PhD, Postdoc

Former Students:

Apurupa Amperayani (PhD Student, Arizona State University)
Jonathan Wells (PhD Student, UT Austin)
Pallavi Atluri
Keerthi Chinthu (Data Scientist, Wabtec Corporation)
Selorm Darkey (Business Intelligent Analyst, Taylor Solutions)
THANK YOU

SBIR: RAE (Realize, Analyze, Engage) - A digital biomarker based detection and intervention system for stress and carvings during recovery from substance abuse disorders.

*Pls: M. Reinhardt, S. Carreiro, P. Indic*

**Department of Veterans Affairs**

*Design of a wearable sensor system and associated algorithm to track suicidal ideation from movement variability and develop a novel objective marker of suicidal ideation and behavior risk in veterans.*

Clinical Science Research and Development Grant (approved for funding).

*P. Indic (site PI, UT-Tyler)*

E.G. Smith (Project PI, VA)

P. Salvatore (Investigator, Harvard University)

**National Science Foundation Smart & Connected Health Grant**

*P. Indic (Lead PI, UT-Tyler)*

D. Paydarfar (Co PI, UT-Austin)

H. Wang (Co PI, UMass Dartmouth)

Y. Kim (Co PI, UMass Dartmouth)

**Pre-Vent**

National Institute Of Health Grant

*P. Indic (Analytical Core PI, UT-Tyler)*

N. Ambal (PI, Univ. of Alabama, Birmingham)

Wearable system for the detection of addiction

*P. Indic (PI, UT-Tyler)*

M. Reinhart (PI, ContinueYou, LLC)

S. Carriero, (PI. Univ. of Mass. Med. School)
DISCUSSION