SIGNAL & DATA ANALYTICS IN IoMT
Tech-in-Med Summer Camp

PREMANANDA INDIC, PH.D.
DEPARTMENT OF ELECTRICAL ENGINEERING

1. Different physiological signals
2. Features of the signals associated with health
3. Differentiating signals and data
5. Development of algorithms
6. Processing of signals
7. Data analytics (Machine Learning)
8. Converting algorithms into software code
9. Embedding the code in the sensors.
What is Machine Learning?

- Machine Learning is a field of study that gives computers the ability to “learn” without being explicitly programmed
  - Prediction
  - Classification

Samuel AL, IBM J. Research & Development, 1959, vol. 3 (3), 210-229
MACHINE LEARNING

➢ Too many books spoil the curiosity

▪ Start with Andrew Ng, Machine Learning, Stanford University available on YouTube

Some Statistics & Programming Knowledge Helps!
MACHINE LEARNING

Analytical Tools

Simple Calculator (Boolean Algebra)

Scientific Calculator (Series Expansion, Boolean Algebra)

Computer (Programming Language, Assembly Language, Series Expansion, Boolean Algebra)

Smart Devices (ML Models, Programming Language, Assembly Language, Series Expansion, Boolean Algebra)
MACHINE LEARNING

➢ Always there is a mathematical foundation

Analytical Tools (Logarithm, Laplace Transform, Fourier Transform…….)
Computational Tools (Boolean Algebra, Taylor Series Expansion,…….)
Programming Languages (Basic, Fortran, C, C++, Java, …….)
Assembly Languages (depending upon the computer processors)
Machine Learning Models
Artificial Intelligence
MACHINE LEARNING

Examples of Smart Systems
- Voice Recognition
- Tumor Detection
- Weather Forecast
- Driverless Cars
MACHINE LEARNING

- Training Data
- Appropriate Model
- Procedure to Train (Make a machine to “learn”)

(Learning Algorithms, Online vs Batch Learning, Instance Based vs Model Based)
- Test Data
MACHINE LEARNING

➢ Machine Learning with MATLAB

https://commons.wikimedia.org/wiki/File:Man_Driving_Car_Cartoon_Vector.svg

http://clipart-library.com/mechanic-cliparts.html

Machine Learning Driving School

You have a complex problem involving a large amount of data and lots of variables. You know that machine learning would be the best approach—but you’ve never used it before. How do you deal with data that’s messy, incomplete, or in a variety of formats? How do you choose the right model for the data?

Sounds daunting? Don’t be discouraged. A systematic workflow will help you get off to a smooth start.
MACHINE LEARNING

- Preprocessing
- Feature Extraction
- Feature Selection

- VARIABLE
- ARTIFACT
- STATISTICAL FEATURES
- NOISE
- SPECTRAL FEATURES
- SIGNAL
- NONLINEAR FEATURES

Data ➔ Preprocessing ➔ Feature Extraction ➔ Feature Selection ➔ Statistical or Machine Learning Models
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:**

\[ \text{MPG} = \text{Intercept} + \text{Weight} + \text{Acceleration} \]

**Estimated Coefficients:**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tStat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>45.155</td>
<td>3.4659</td>
<td>13.028</td>
<td>1.6266e-22</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.0082475</td>
<td>0.00059836</td>
<td>-13.783</td>
<td>5.3165e-24</td>
</tr>
<tr>
<td>Acceleration</td>
<td>0.19094</td>
<td>0.14743</td>
<td>1.3359</td>
<td>0.18493</td>
</tr>
</tbody>
</table>

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

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

VARIABLE

Real Estate Data

```
Command Window

>> lm=fitlm(housing)

lm =

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

Estimated Coefficients:

```
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tStat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.585e+06</td>
<td>62901</td>
<td>-57.001</td>
<td>0</td>
</tr>
<tr>
<td>longitude</td>
<td>-42730</td>
<td>717.09</td>
<td>-59.588</td>
<td>0</td>
</tr>
<tr>
<td>latitude</td>
<td>-42510</td>
<td>676.95</td>
<td>-62.796</td>
<td>0</td>
</tr>
<tr>
<td>housing_median_age</td>
<td>1157.9</td>
<td>43.389</td>
<td>26.687</td>
<td>2.946e-154</td>
</tr>
<tr>
<td>total_rooms</td>
<td>-8.2497</td>
<td>0.79426</td>
<td>-10.387</td>
<td>3.2948e-25</td>
</tr>
<tr>
<td>total_bedrooms</td>
<td>113.82</td>
<td>6.9306</td>
<td>16.423</td>
<td>3.1889e-60</td>
</tr>
<tr>
<td>population</td>
<td>-38.386</td>
<td>1.0841</td>
<td>-35.407</td>
<td>1.4597e-266</td>
</tr>
<tr>
<td>households</td>
<td>47.701</td>
<td>7.5466</td>
<td>6.3209</td>
<td>2.6535e-10</td>
</tr>
<tr>
<td>median_income</td>
<td>40298</td>
<td>337.21</td>
<td>119.5</td>
<td>0</td>
</tr>
</tbody>
</table>
```
DIFFERENT TYPES OF FEATURES

➢ VARIABLE

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

STATISTICAL FEATURES

Individual with bipolar disorder

Mean

Kurtosis

N = 128
DIFFERENT TYPES OF FEATURES

STATISTICAL FEATURES

Correlation with Self Reported Suicidal Ideation

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

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

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

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

MATLAB functions

mean(filename)

variance(filename)

skewness(filename)

kurtosis(filename)

\([r,p]=\text{corr(resultsfilename(:,1),resultsfilename(:,5))};\)
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, see the `Kmax` 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, see the `N` 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, 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 `Prominence` name-value pair argument.
DIFFERENT TYPES OF FEATURES

➢ SPECTRAL FEATURES

Wavelet transform

```python
wavelets(filename)
```

DIFFERENT TYPES OF FEATURES

SPECTRAL FEATURES

Individuals during major depression phase

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

Statistical vs. Machine Learning Models

- VARIABLE
- ARTIFACT
- STATISTICAL FEATURES
- NOISE
- SPECTRAL FEATURES
- SIGNAL
- NONLINEAR FEATURES

Data ➔ Preprocessing ➔ Feature Extraction ➔ Feature Selection ➔ Statistical or Machine Learning Models
Statistical vs. Machine Learning Models

Purpose:

Statistical models are used for inference (To find association between features and an outcome). Results should be interpretable.

Machine Learning models are used for prediction (Use features that can predict an outcome). Results may not be interpretable.
Statistical vs. Machine Learning Models

Association vs. Prediction

Healthy Individual

Individual with depression

\[ V_I = m \times S_I + C \]
\[ m = r \frac{\sigma_{VI}}{\sigma_{SI}} \]
\[ C = \mu_{VI} - m\mu_{SI} \]

\[ \tilde{S}_I = a \times V_I + b \]

Sensitivity & Specificity
LEARNING APPROACHES

➢ Supervised Learning
   Learning a relationship between features and the outcome using a training set

➢ Unsupervised Learning
   Learning underlying structures in features
LEARNING APPROACHES

➢ Supervised Learning
  • Linear Regression
  • Logistic Regression
  • Support Vector Machine
  • Artificial Neural Network
  • ................
  • ................
  • ................
LEARNING APPROACHES

➢ Unsupervised Learning

  Clustering

  • Principal Component Analysis
  • Independent Component Analysis
  • Singular Value Decomposition
  • .......
  • ........
LEARNING APPROACHES

➢ Do machines actually “learn”? 

\[ VI = m \times SI + C \]
LEARNING APPROACHES

Do machines actually “learn”?

\[ e(N = 1) = \overline{VI}(N = 1) - VI(N = 1) \]
\[ e(N = 2) = \overline{VI}(N = 2) - VI(N = 2) \]

\[ e(N = 24) = \overline{VI}(N = 24) - VI(N = 24) \]

\[ E = \sum_{n=1}^{N} e^2 \]

\[ \overline{VI} = m \times SI + C \]
LEARNING APPROACHES

➢ Do machines actually “learn”?

How do we find minimum E?

\[
\text{\( \bar{V}I = m \times SI + C \)}
\]

<table>
<thead>
<tr>
<th>( m )</th>
<th>( 0.1 )</th>
<th>( 0.6 )</th>
<th>( 0.8 )</th>
<th>( 0.01 )</th>
<th>( 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C )</td>
<td>0.1</td>
<td>0.6</td>
<td>0.8</td>
<td>0.01</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>0.01</td>
<td>0.001</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>0.0006</td>
<td>0.03</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>12</td>
<td>0.1</td>
<td>12</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0.5</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

\[ V_{\text{I}} = m \times SI + C \]
LEARNING APPROACHES

➢ Do machines actually “learn”?

How do we find minimum E?

- Gradient Descent

by Louis Augustin Cauchy in 1847

\[
\bar{VI} = m \times SI + C
\]

Linear Regression

\[
\hat{SI} = a \times VI + b
\]
LEARNING APPROACHES

➢ Do machines actually “learn”? 

Classification of High Risk (n=43) vs. Low Risk (n=95) 

0 = Low Risk, 1 = High Risk

\[ p = \frac{1}{1 + e^{-\alpha VI - \beta}} \]

Mean 
Variance 
Skewness 
Kurtosis 
Power 
Period

Linear Regression

Logistic Regression

Accuracy ~73%
LEARNING APPROACHES

➢ How to implement in MATLAB?

**Step 1:** Create an excel sheet with features with class assignments
LEARNING APPROACHES

➢ How to implement in MATLAB?

**Step 2:** Open MATLAB and drag the excel file to workspace.
LEARNING APPROACHES

➢ How to implement in MATLAB?

Step 3: Click Import Selection and import data
LEARNING APPROACHES

➢ How to implement in MATLAB?

**Step 4:** Features are in workspace and ready
How to implement in MATLAB?

**Step 5:**
- Go to Apps,
- click classification learner,
- select Logistic Regression from Model Type
- click New Session,
- select from Workspace
LEARNING APPROACHES

➢ How to implement in MATLAB?

**Step 6:** Set 10 fold Cross validation
- Start the session
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING

➢ UNSUPERVISED LEARNING
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING

➢ UNSUPERVISED LEARNING
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING (Classification / Prediction)

Provide training set with features and solutions
SUPERVISED LEARNING (Classification / Prediction)

Find the area of a rectangle

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12.1</td>
<td>13.4</td>
<td>162.3</td>
<td>25.5</td>
<td>-1.3</td>
<td>162.14</td>
<td>0.90</td>
</tr>
<tr>
<td>8.6</td>
<td>9.7</td>
<td>83.4</td>
<td>18.3</td>
<td>-1.1</td>
<td>83.42</td>
<td>0.89</td>
</tr>
<tr>
<td>3.2</td>
<td>5.4</td>
<td>17.3</td>
<td>8.6</td>
<td>-2.2</td>
<td>17.28</td>
<td>0.59</td>
</tr>
<tr>
<td>6.1</td>
<td>10.2</td>
<td>62.25</td>
<td>16.3</td>
<td>-4.1</td>
<td>62.22</td>
<td>0.60</td>
</tr>
<tr>
<td>18.2</td>
<td>6.4</td>
<td>116.5</td>
<td>24.6</td>
<td>11.8</td>
<td>116.48</td>
<td>2.83</td>
</tr>
<tr>
<td>1.6</td>
<td>2.8</td>
<td>4.5</td>
<td>4.4</td>
<td>-1.2</td>
<td>4.48</td>
<td>0.57</td>
</tr>
<tr>
<td>7.7</td>
<td>0.6</td>
<td>4.7</td>
<td>8.3</td>
<td>7.1</td>
<td>4.62</td>
<td>12.83</td>
</tr>
</tbody>
</table>
## APPROACHES IN DETAIL

### SUPERVISED LEARNING (Classification / Prediction)

Find the area of a rectangle

<table>
<thead>
<tr>
<th>L</th>
<th>W</th>
<th>A</th>
<th>E1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.1</td>
<td>13.4</td>
<td>162.3</td>
<td>136.8</td>
<td>163.6</td>
<td>0.16</td>
<td>161.40</td>
</tr>
<tr>
<td>8.6</td>
<td>9.7</td>
<td>83.4</td>
<td>65.1</td>
<td>84.5</td>
<td>0.02</td>
<td>82.51</td>
</tr>
<tr>
<td>3.2</td>
<td>5.4</td>
<td>17.3</td>
<td>8.7</td>
<td>19.5</td>
<td>0.02</td>
<td>16.71</td>
</tr>
<tr>
<td>6.1</td>
<td>10.2</td>
<td>62.25</td>
<td>45.95</td>
<td>66.35</td>
<td>0.03</td>
<td>61.65</td>
</tr>
<tr>
<td>18.2</td>
<td>6.4</td>
<td>116.5</td>
<td>91.90</td>
<td>104.70</td>
<td>0.02</td>
<td>113.66</td>
</tr>
<tr>
<td>1.6</td>
<td>2.8</td>
<td>4.5</td>
<td>0.1</td>
<td>5.7</td>
<td>0.02</td>
<td>3.93</td>
</tr>
<tr>
<td>7.7</td>
<td>0.6</td>
<td>4.7</td>
<td>3.6</td>
<td>2.4</td>
<td>0.08</td>
<td>8.13</td>
</tr>
</tbody>
</table>
APPROACHES IN DETAIL

SUPERVISED LEARNING (Classification / Prediction)

- Linear Regression
- Logistic Regression
- Support Vector Machines
- k-Nearest Neighbors
- Decision Trees and Random Forests
- Neural Networks
APPROACHES IN DETAIL

- SUPERVISED LEARNING (Classification / Prediction)
  - Linear Regression

Given $m$ outcomes $y^i$ where $i = 1, 2, \ldots, m$ with each outcome depends on $n$ features $x_j$ where $j = 1, 2, \ldots, n$. Find the best estimate of $y^i$ as $\hat{y}^i$ using the $n$ features with appropriate parameters $\theta_j$ such that $J = \left\langle (\hat{y}^i - y^i)^2 \right\rangle$

$$\hat{y}^i = \theta_0^i + \theta_1^i x_1^i + \theta_2^i x_2^i + \cdots + \theta_n^i x_n^i$$
APPROACHES IN DETAIL

SUPervised Learning (Classification / Prediction)

- Linear Regression

\[
\hat{y}^i = \theta_0^i + \theta_1^i x_1^i + \theta_2^i x_2^i + \cdots + \theta_n^i x_n^i
\]

\[
\hat{y} = \Theta X = h_\theta(X)
\]

Cost Function to Minimize

\[
J = \left( (\hat{y}^i - y^i)^2 \right) = (\hat{Y} - Y)^T (\hat{Y} - Y)
\]
APPROACHES IN DETAIL

SUPERVISED LEARNING (Classification / Prediction)

- Linear Regression

\[ \hat{y}_i = \theta_0^i + \theta_1^i x_1^i + \theta_2^i x_2^i + \cdots + \theta_n^i x_n^i \]

\[ \hat{Y} = \Theta X = h_\theta(X) \]

- Gradient Descent by Louis Augustin Cauchy in 1847

Cost Function to Minimize

\[ J = \left( (\hat{y}_i - y_i)^2 \right) = (\hat{Y} - Y)^T (\hat{Y} - Y) \]
SUPervised Learning (Classification / Prediction)

- Logistic Regression

\[
\hat{p} = f(\Theta.X) = h_\Theta(X)
\]

\[
\hat{y} = 1 \text{ if } \hat{p} < 0.5; \quad \hat{y} = 0 \text{ if } \hat{p} \geq 0.5
\]

Derive Cost Function to Minimize \( J \)
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING (Classification / Prediction)
  • Linear Regression
    Mainly for regression (predicting an outcome)
  • Logistic Regression
    Mainly for classification (0 or 1)

High Risk vs. Low Risk

https://medium.datadriveninvestor.com/machine-learning-101-part-1-24835333d38a
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING (Classification / Prediction)
  • Support Vector Machine

Used for regression as well as classification
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING (Classification / Prediction)
  • Support Vector Machine (SVM)

Used for regression as well as classification

---

Linear vs. SVM Regression

SVM

Linear

\[ x \]

\[ y \]

\[ x \]
APPROACHES

➢ SUPERVISED LEARNING (Classification / Prediction)
  • Support Vector Machine (SVM)

Used for regression as well as classification

https://medium.com/@LSchultebraucks/introduction-to-support-vector-machines-9f8161ae2fcb
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING (Classification / Prediction)
  • Support Vector Machine (SVM)

Used for regression as well as classification

https://www.mathworks.com/matlabcentral/fileexchange/62061-multi-class-svm
APPROACHES IN DETAIL

➢ SUPERVISED LEARNING (Classification / Prediction)
  • Linear Regression
  • Logistic Regression
  • Support Vector Machines
  • k-Nearest Neighbors
  • Decision Trees and Random Forests
  • Neural Networks
Project 1

- Prediction of House Price (housing.csv) Regression Problem

longitude
latitude
housing_median_age
total_rooms
total_bedrooms
population
households median_income
median_house_value
ocean_proximity
Project 2

Prediction of House Price (housing.csv) Classification Problem

longitude
latitude
housing_median_age
total_rooms
total_bedrooms
population
households median_income
median_house_value (High/Low) Threshold= 257500
ocean_proximity
Project 2

Prediction of House Price (housing.csv) Classification Problem

Confusion Matrix

<table>
<thead>
<tr>
<th>True Class</th>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

True Positive = True Positive / Total Positive

True Negative Rate = True Negative / Total Negative = 1 – False Positive Rate
Project 2

➢ Prediction of House Price (housing.csv)  Classification Problem
Project 3

➢ To test the hypothesis that the features of SpO2 can detect smoker from non-smoker
Project 4

➢ To test the hypothesis that the features of saliva can detect COPD from other conditions

➢ Data set: http://archive.ics.uci.edu/ml/datasets/Exasens

https://cml.ics.uci.edu/