MACHINE LEARNING: REGRESSION

PREMANANDA INDIC, PH.D.

DEPARTMENT OF ELECTRICAL ENGINEERING
ANALYSIS PLATFORM

University of Texas at Tyler

MATLAB Access for Everyone at

University of Texas at Tyler

ANALYSIS PLATFORM

University of Texas at Tyler

MATLAB Access for Everyone at

University of Texas at Tyler

OUTLINE

- INTRODUCTION
- DIFFERENT REGRESSION APPROACHES
- EXAMPLES
OUTLINE

- INTRODUCTION

- DIFFERENT REGRESSION APPROACHES

- EXAMPLES
INTRODUCTION

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
INTRODUCTION

What is Machine Learning?

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

Samuel Al, IBM J. Research & Development, 1959, vol. 3 (3), 210-229
OUTLINE

- INTRODUCTION
- DIFFERENT REGRESSION APPROACHES
- EXAMPLES
APPROACHES

- SUPERVISED LEARNING
- UNSUPERVISED LEARNING
APPROACHES

 SUPERVISED LEARNING (Classification / Prediction)

Provide training set with features and solutions
APPROACHES

- STANDARD MACHINE LEARNING
- ADVANCED MACHINE LEARNING

Based on Artificial Neural Networks (Deep Learning)
APPROACHES

REGRESSION

• Linear Regression
• Support Vector Regression
APPROACHES

REGRESSION

• Linear Regression

• Support Vector Regression
APPROACHES

- 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 $J = \left\langle (\hat{y}^{(i)} - y^{(i)})^2 \right\rangle$

$$\hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \ldots + \theta_n x_n^{(i)}$$
APPROACHES

- Linear Regression

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

\[ \hat{Y} = \Theta^T X \]

\[ \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & \ldots & x_1^{(m)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & x_n^{(3)} & \ldots & x_n^{(m)} \end{bmatrix} \]

Cost Function to Minimize

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

- **Linear Regression**

  \[
  J = \left(\hat{y}^i - y^i\right)^2 = (\hat{Y} - Y)^T (\hat{Y} - Y) = (\Theta^T X - Y)^T (\Theta^T X - Y)
  \]

  \[
  \frac{dJ}{d\Theta} = 0
  \]

  \[
  \Theta = (X^T X)^{-1} X^T Y
  \]
Linear Regression

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

\[ \hat{Y} = \theta^T X \]

- Gradient Descent by Louis Augustin Cauchy in 1847

Cost Function to Minimize

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

- Linear Regression

$$\theta^{k+1} = \theta^k - \gamma \nabla_{\theta} J(\theta)$$

$$\nabla_{\theta} J(\theta) = \frac{2}{m} X^T (X\theta - Y)$$
APPROACHES

- Polynomial 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 $J = \left\{ (\hat{y}^{(i)} - y^{(i)})^2 \right\}$

\[
\hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_1^{2(i)} + \cdots + \theta_n x_1^{n(i)}
\]
APPROACHES

 Polynomial Regression

\[ \hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_1^{2(i)} + \cdots + \theta_n x_1^{n(i)} \]

\[ \hat{Y} = \Theta^T X \quad \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \cdots \\ \theta_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & \cdots & x_1^{(m)} \\ x_1^{2(1)} & x_1^{2(2)} & x_1^{2(3)} & \cdots & x_1^{2(m)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_1^{n(1)} & x_1^{n(2)} & x_1^{n(3)} & \cdots & x_1^{n(m)} \end{bmatrix} \]

Cost Function to Minimize

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

REGRESSION

• Linear Regression

• Support Vector Regression
APPROACHES

- Support Vector Regression

\[-\varepsilon < y - f(x) < \varepsilon\]

\[f(x) = \theta_0 + \theta x \text{ (Linear Regression)}\]

\[f(x) = \theta_0 + \sum_{i=1}^{m} G(x^i, x)\]

\[G(x^i, x) = x^i \cdot x \text{ (Linear SVR)}\]

\[G(x_j, x_k) = \exp\left(-\|x_j - x_k\|^2\right)\]

\[G(x_j, x_k) = (1 + x_j'x_k)^q, \text{ where } q \text{ is in the set \{2,3,\ldots\}.}\]
EXAMPLE 1

Home Value Prediction (App Based): 9 features to predict medianHouseValue (N=20640)

- **longitude**: A measure of how far west a house is; a higher value is farther west
- **latitude**: A measure of how far north a house is; a higher value is farther north
- **housingMedianAge**: Median age of a house within a block; a lower number is a newer building
- **totalRooms**: Total number of rooms within a block
- **totalBedrooms**: Total number of bedrooms within a block
- **population**: Total number of people residing within a block
- **households**: Total number of households, a group of people residing within a home unit, for a block
- **medianIncome**: Median income for households within a block of houses (measured in tens of thousands of US Dollars)
- **medianHouseValue**: Median house value for households within a block (measured in US Dollars)

oceanProximity: Location of the house w.r.t ocean/sea

Demo with N=5000
70% Training Data
30% Test Data
Models Trained:
Linear Regression
SVM

https://www.kaggle.com/camnugent/california-housing-prices
EXAMPLE 1

- Home Value Prediction (App Based): 9 features to predict medianHouseValue (N=5000)

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Validation (10 fold) RMSE</th>
<th>R-squared</th>
<th>Test RMSE</th>
<th>Test R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression (using App)</td>
<td>69010</td>
<td>0.64</td>
<td>65501</td>
<td>0.67</td>
</tr>
<tr>
<td>Linear SVM (using App)</td>
<td>70382</td>
<td>0.64</td>
<td>66858</td>
<td>0.66</td>
</tr>
</tbody>
</table>
EXAMPLE 1

- Home Value Prediction (App Based): 9 features to predict medianHouseValue (N=5000)
EXAMPLE 2

Home Value Prediction (Realistic Approach): 9 features to predict medianHouseValue (N=5000)

1. Visualize the data
2. Identify the features (find correlations between variables)
3. Preprocess the data (missing values, outliers)
4. Train the Model
5. Select the best performance model
EXAMPLE 2

Home Value Prediction (Realistic Approach): 9 features to predict medianHouseValue (N=5000)

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Validation RMSE</th>
<th>Test RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin regression</td>
<td>70071</td>
<td>65501</td>
</tr>
<tr>
<td>Lin. Regression – fewer variables</td>
<td>69031</td>
<td>65357</td>
</tr>
<tr>
<td>SVM – linear kernel</td>
<td>116370</td>
<td>116130</td>
</tr>
<tr>
<td>SVM – Gaussian Kernel</td>
<td>60099</td>
<td>57708</td>
</tr>
</tbody>
</table>
LASSO REGRESSION

- Linear Regression

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

\[ \hat{Y} = \Theta^T X \]

- Gradient Descent by **Louis Augustin Cauchy** in 1847

Cost Function to Minimize

\[ J = \left\{ \left( \hat{y}^i - y^i \right)^2 \right\} = (\hat{Y} - Y)^T (\hat{Y} - Y) \]
LASSO REGRESSION

- Linear Regression with Lasso

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

\[ \hat{Y} = \Theta^T X \]

Cost Function to Minimize

\[ J = \left( (\hat{y}^i - y^i)^2 \right) = (\hat{Y} - Y)^T (\hat{Y} - Y) + \lambda \sum_{j=1}^{n} |\theta_j| \]
EXAMPLE 3

Home Value Prediction (Lasso Regression): 9 features to predict medianHouseValue (N=5000)

\[ J = \mathbb{E}[(\hat{y} - y)^2] = (\hat{Y} - Y)^T(\hat{Y} - Y) + \lambda \sum_{j=1}^{n}|\theta_j| \]

Lambda

Lasso removes the ‘total_rooms’ and ‘Ocean Proximity_inland’ variables as least important.

RMSE on test data with 7 features = 66443
CONCLUSION

- Regression provides continuous prediction of an outcome with selected features.
- Understanding of features in relation to outcome is important.
- Several codes are available to perform regression analysis.
THANK YOU

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

Pis: M. Reinhardt, S. Carreiro, P. Indic

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.

Department of Veterans Affairs

P. Indic (site PI, UT-Tyler)
E.G. Smith (Project PI, VA)
P. Salvatore (Investigator, Harvard University)

Design of a wearable biosensor sensor system with wireless network for the remote detection of life threatening events in neonates.

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)

NSTF

The University of Texas System

P. Indic (PI, UT Tyler)

Pre-Vent

National Institute Of Health Grant

P. Indic (Analytical Core PI, UT-Tyler)
N. Ambal (PI, Univ. of Alabama, Birmingham)

ORS Research Design & Data Analysis Lab

Office of Research and Scholarship

STARs Award

ViSiOn

P. Indic (site PI, UT-Tyler)
P. Ramanand (Co-I, UT Tyler)
N. Ambal, (PI, Univ. of Alabama, Birmingham)

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

Pis: M. Reinhardt, S. Carreiro, P. Indic

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.

Department of Veterans Affairs

P. Indic (site PI, UT-Tyler)
E.G. Smith (Project PI, VA)
P. Salvatore (Investigator, Harvard University)

Design of a wearable biosensor sensor system with wireless network for the remote detection of life threatening events in neonates.

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)

NSTF

The University of Texas System

P. Indic (PI, UT Tyler)

Pre-Vent

National Institute Of Health Grant

P. Indic (Analytical Core PI, UT-Tyler)
N. Ambal (PI, Univ. of Alabama, Birmingham)

ORS Research Design & Data Analysis Lab

Office of Research and Scholarship

STARs Award

ViSiOn

P. Indic (site PI, UT-Tyler)
P. Ramanand (Co-I, UT Tyler)
N. Ambal, (PI, Univ. of Alabama, Birmingham)
QUESTIONS