MACHINE LEARNING: REGRESSION

PREMANANDA INDIC, PH.D.

DEPARTMENT OF ELECTRICAL ENGINEERING

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

ORS Research Design & Data Analysis Lab

Office of Research and Scholarship

ANALYSIS PLATFORM

MathWorks®

University of Texas at Tyler

Get Software | Learn MATLAB | Teach with MATLAB | What's New

MATLAB Access for Everyone at

University of Texas at Tyler

https://www.mathworks.com/academia/tah-portal/university-of-texas-at-tyler-1108545.html

ANALYSIS PLATFORM

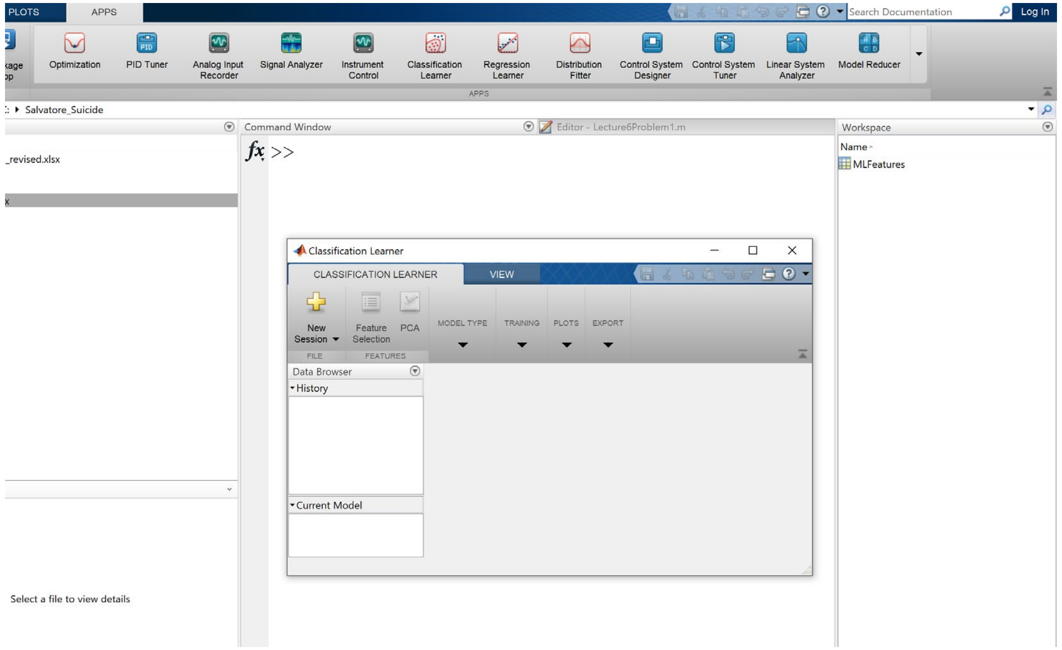
MathWorks®

University of Texas at Tyler

Get Software | Learn MATLAB | Teach with MATLAB | What's New

MATLAB Access for Everyone at

University of Texas at Tyler



https://www.mathworks.com/academia/tah-portal/university-of-texas-at-tyler-1108545.html

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

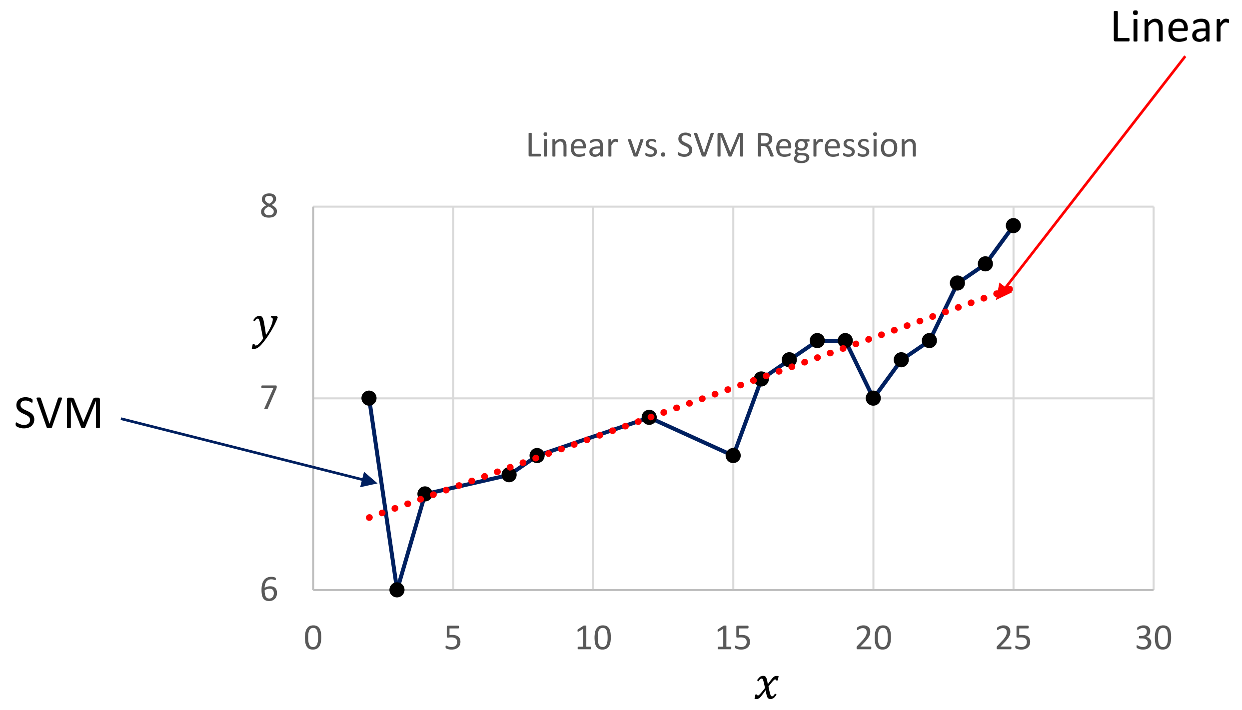
- ADVANCED MACHINE LEARNING

  Based on Artificial Neural Networks (Deep Learning)

APPROACHES

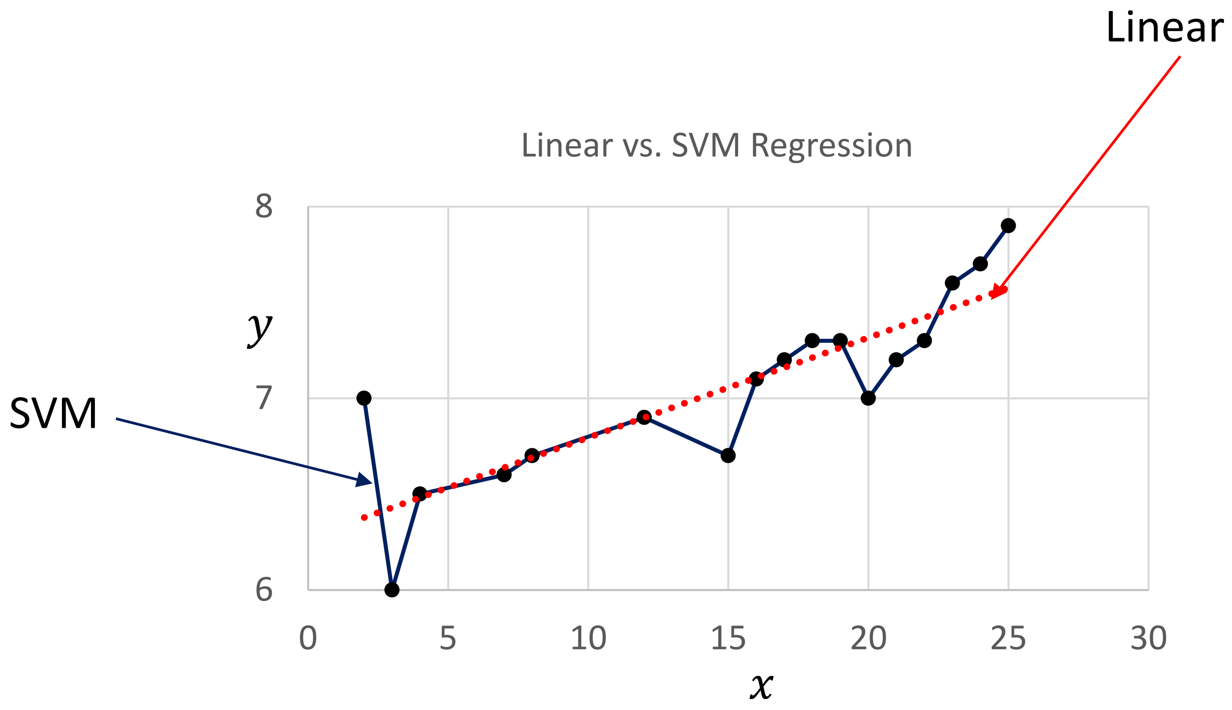
➢ REGRESSION

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

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

APPROACHES

- Linear Regression

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

$$\hat{Y} = \Theta^T X \qquad \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ .. \\ .. \\ \theta_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 1 & .. & .. & 1 \\ x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & .. & .. & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & x_2^{(3)} & .. & .. & x_2^{(m)} \\ .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. \\ x_n^{(1)} & x_n^{(2)} & x_n^{(3)} & .. & .. & x_n^{(m)} \end{bmatrix}$$

Cost Function to Minimize

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

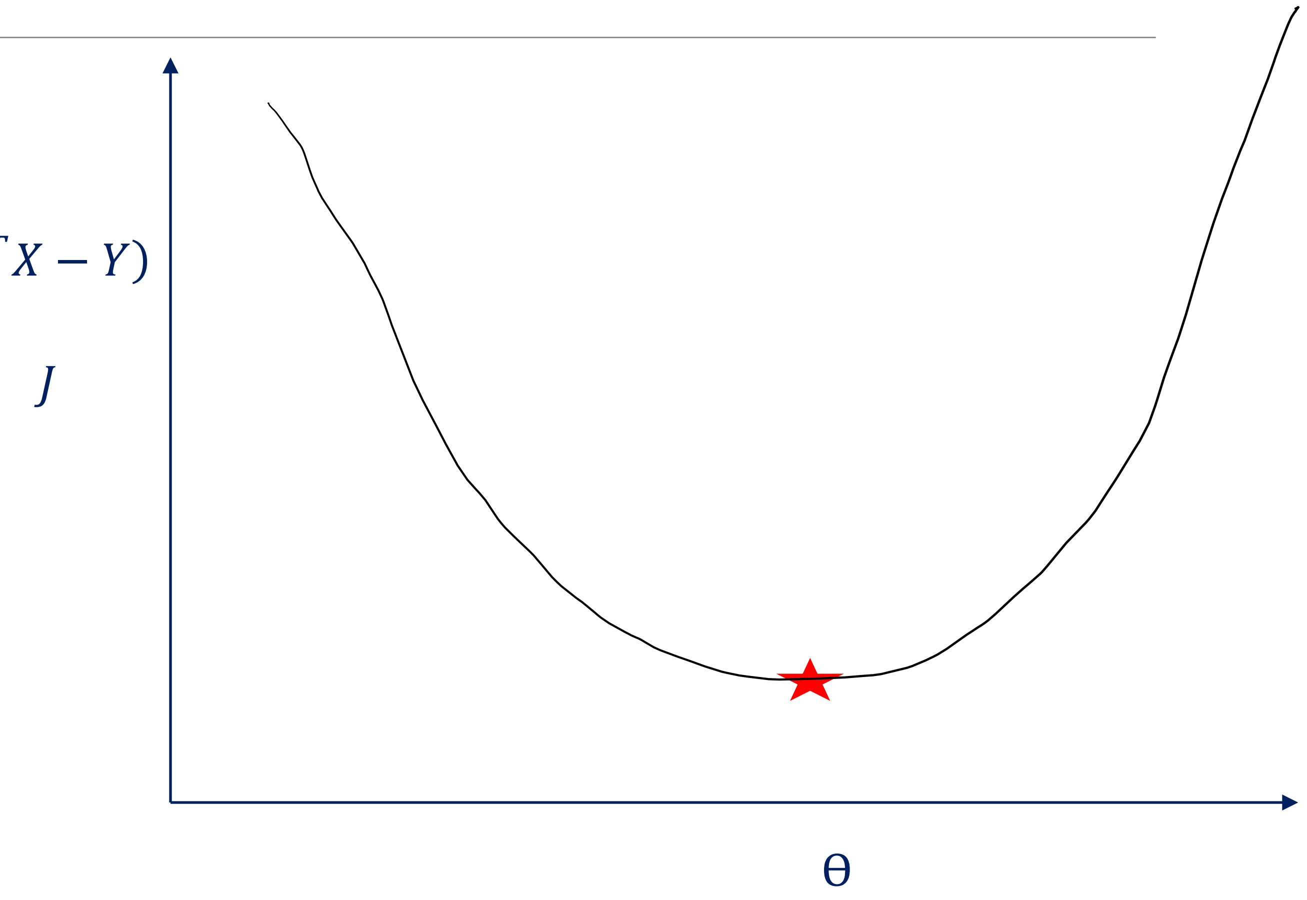
APPROACHES

- Linear Regression

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

$$\frac{dJ}{d\Theta} = 0$$

$$\Theta = (X^T X)^{-1} X^T Y$$



APPROACHES

- Linear Regression

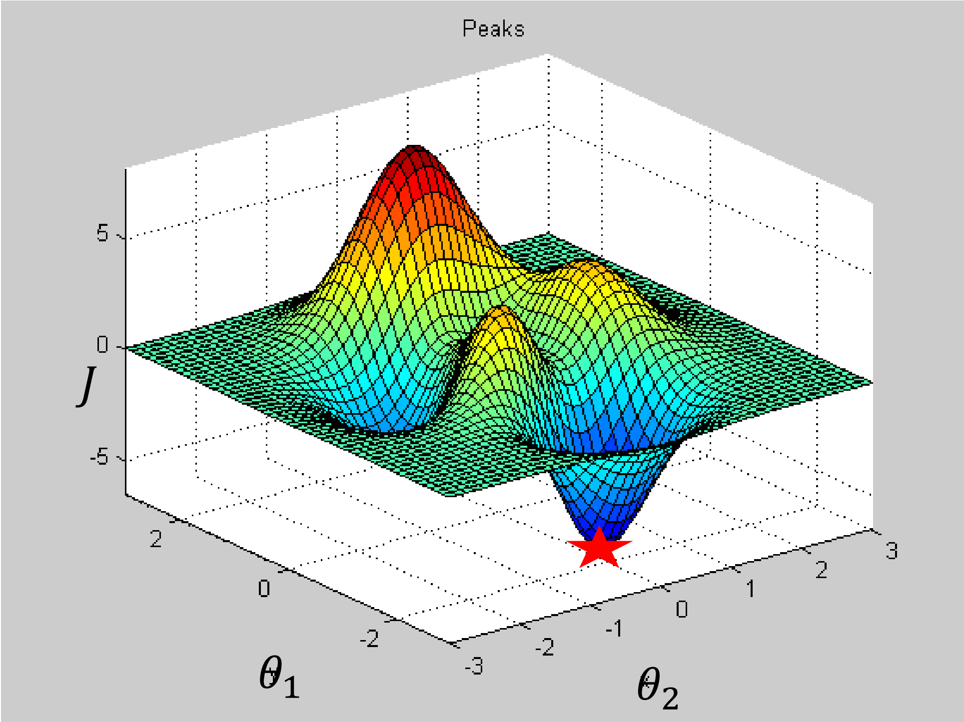
$$\hat{y}^i = \theta_0 + \theta_1 x_1^i + \theta_2 x_2^i + \cdots \ldots \ldots . + \theta_n x_n^i$$

$$\hat{Y} = \Theta^T X$$

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

Cost Function to Minimize

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

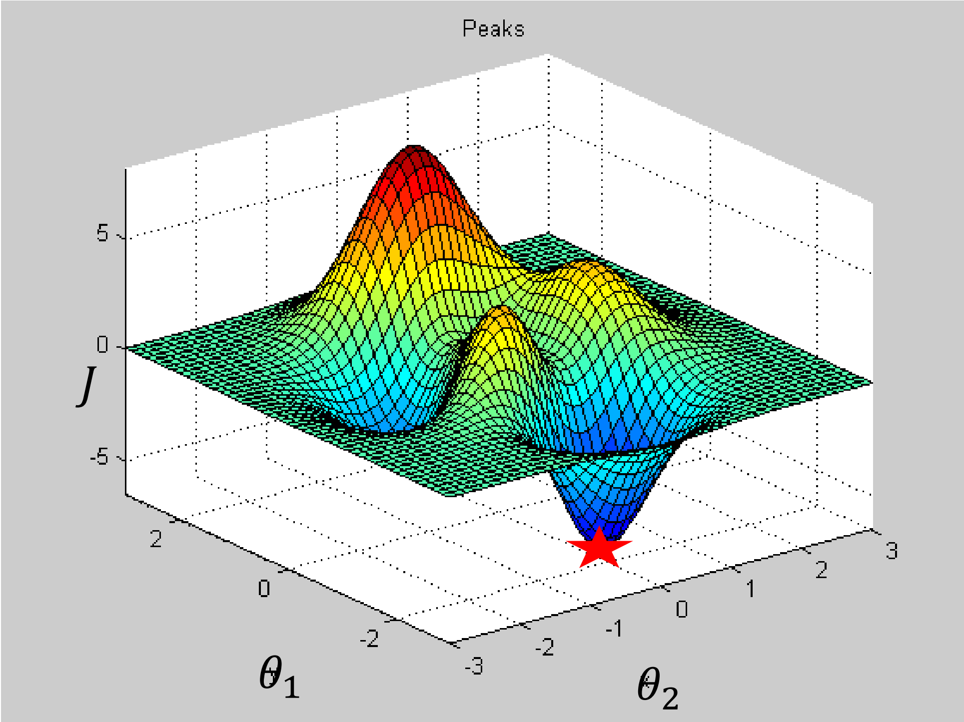


APPROACHES

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

$$\hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_1^{2(i)} + \cdots \ldots \ldots . + \theta_n x_1^{n(i)}$$

APPROACHES

- Polynomial Regression

$$\hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_1^{2(i)} + \cdots \dots \dots . + \theta_n x_1^{n(i)}$$

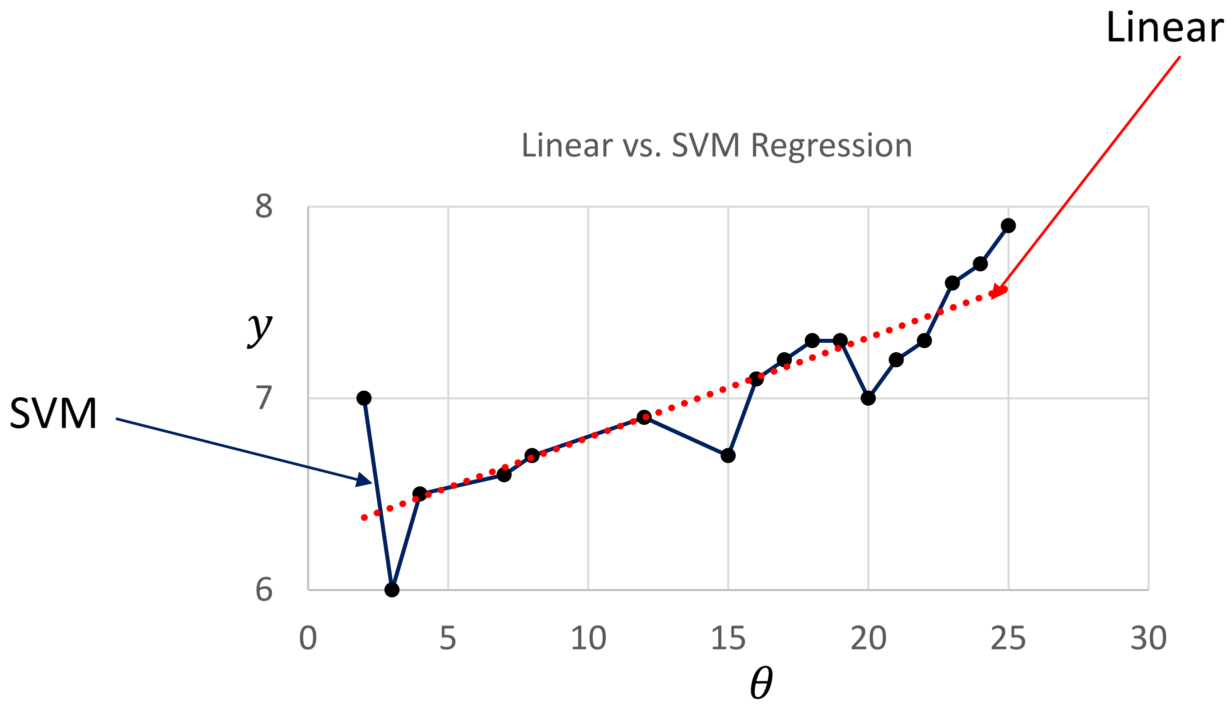
$$\hat{Y} = \Theta^T X \qquad \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ .. \\ .. \\ \theta_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 1 & .. & .. & 1 \\ x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & .. & .. & x_1^{(m)} \\ x_1^{2(1)} & x_1^{2(2)} & x_1^{2(3)} & .. & .. & x_1^{2(m)} \\ .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. \\ x_1^{n(1)} & x_1^{n(2)} & x_1^{n(3)} & .. & .. & x_1^{n(m)} \end{bmatrix}$$

Cost Function to Minimize

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

APPROACHES

- REGRESSION
  - Linear Regression
  - Support Vector Regression



APPROACHES

- Support Vector Regression

$$-\epsilon < y - f(x) < \epsilon$$

$f(x) = \theta_0 + \theta x$ (Linear Regression)

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

$G(x^i, x) = x^i . x$ (Linear SVR)

$G(x_j, x_k) = \exp\left(-\|x_j - x_k\|^2\right)$
$G(x_j, x_k) = (1 + x_j{}' x_k)^q$, where q is in the set {2,3,...}.

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

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

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

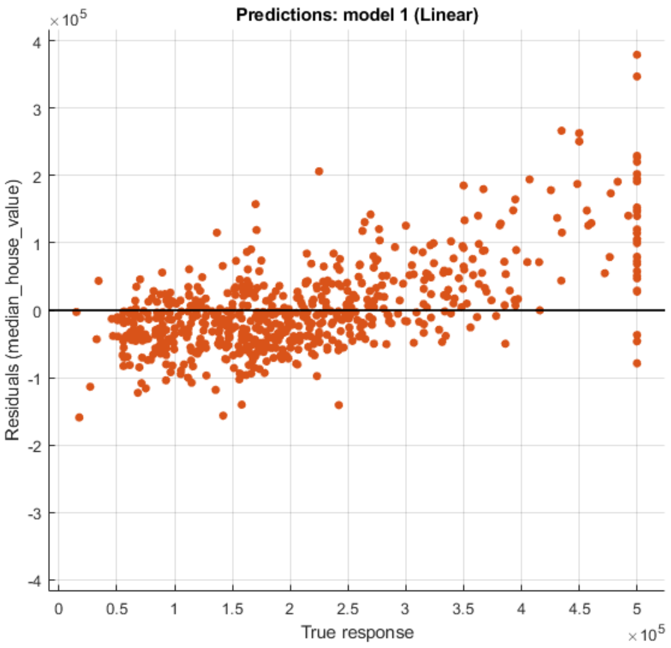
EXAMPLE 1

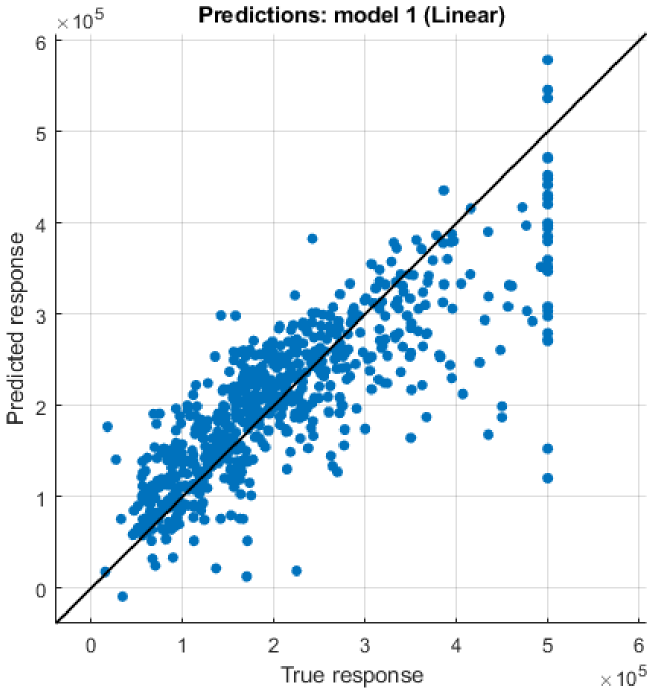
- Home Value Prediction (App Based): 9 features to predict medianHouseValue (N=5000)

Model Type	Validation (10 fold) RMSE	R-squared	Test RMSE	Test R-squared
Linear Regression (using App)	69010	0.64	65501	0.67
Linear SVM (using App)	70382	0.64	66858	0.66

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)

Model Type	Validation RMSE	Test RMSE
Lin regression	70071	65501
Lin. Regression – fewer variables	69031	65357
SVM –linear kernel	116370	116130
SVM –Gaussian Kernel	60099	57708

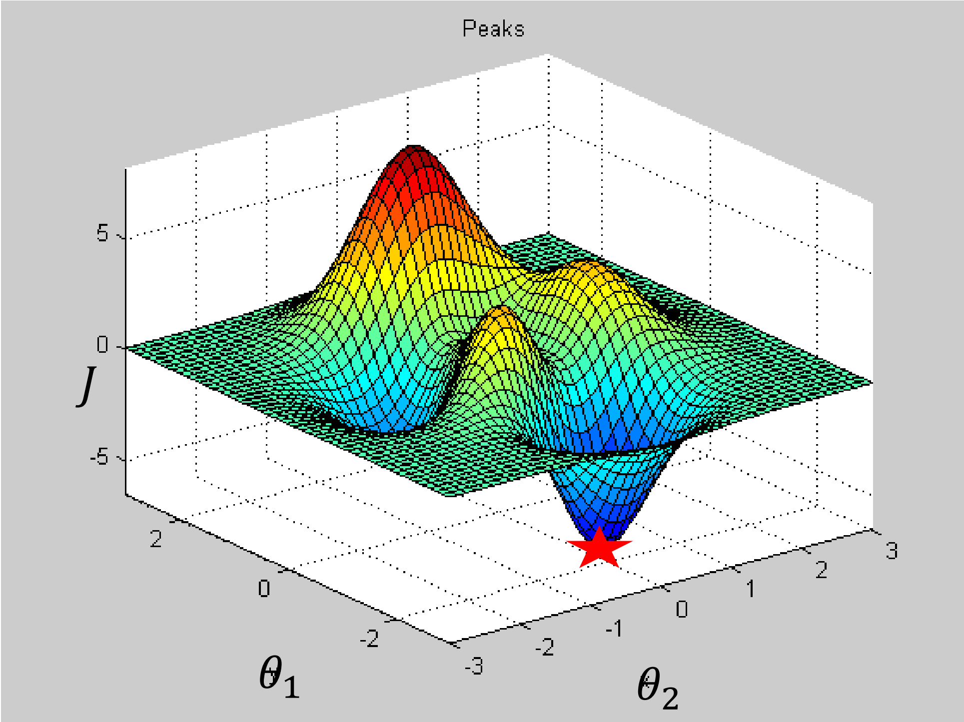
LASSO RGRESSION

- Linear Regression

$$\hat{y}^i = \theta_0 + \theta_1 x_1^i + \theta_2 x_2^i + \cdots \dots \dots . + \theta_n x_n^i$$

$$\hat{Y} = \Theta^T X$$

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



Cost Function to Minimize

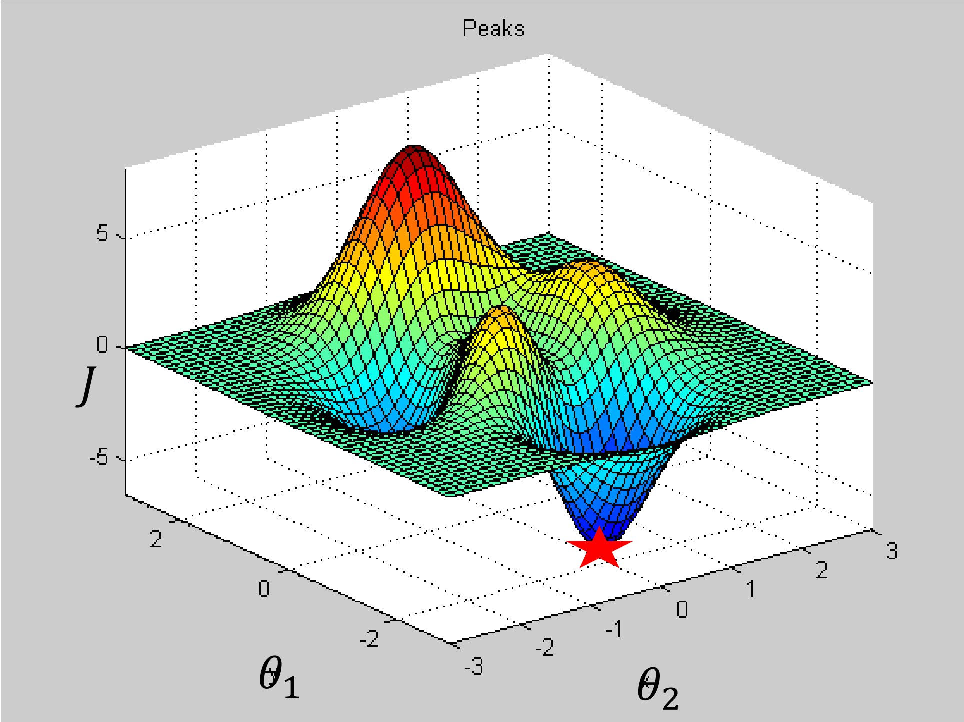
$$J = \left\langle \left(\hat{y}^i - y^i\right)^2 \right\rangle = (\hat{Y} - Y)^T (\hat{Y} - Y)$$

LASSO RGRESSION

- Linear Regression with Lasso

$$\hat{y}^i = \theta_0 + \theta_1 x_1^i + \theta_2 x_2^i + \cdots \ldots \ldots . + \theta_n x_n^i$$

$$\hat{Y} = \Theta^T X$$

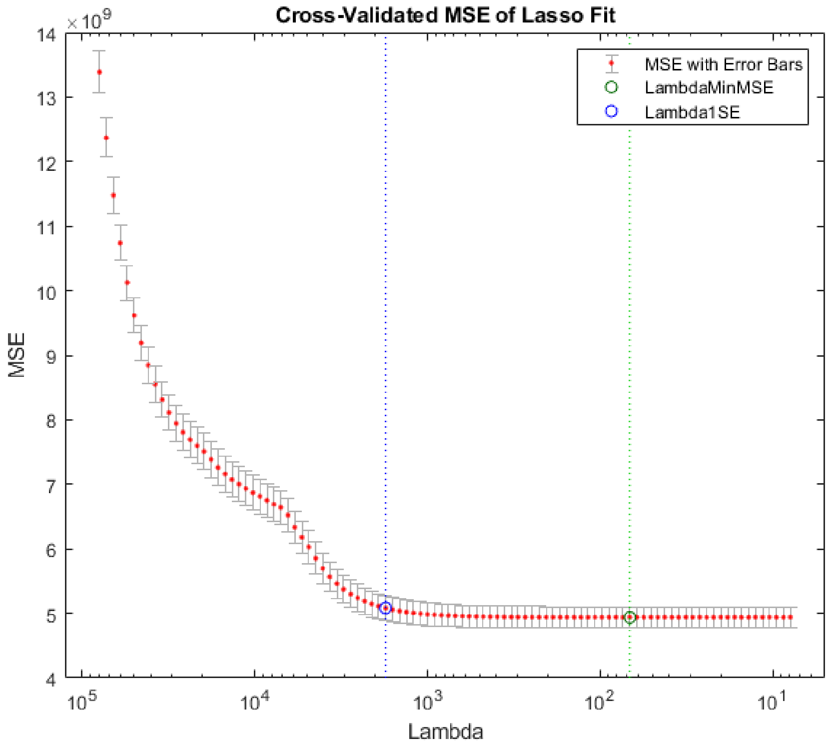


Cost Function to Minimize

$$J = \left\langle \left(\hat{y}^i - y^i\right)^2 \right\rangle = \left(\hat{Y} - Y\right)^T \left(\hat{Y} - Y\right) + \lambda \sum_{j=1}^{n} |\theta_j|$$

EXAMPLE 3

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



$$J = \left\langle \left(\hat{y}^i - y^i\right)^2 \right\rangle = \left(\hat{Y} - Y\right)^T \left(\hat{Y} - Y\right) + \lambda \sum_{j=1}^{n} |\theta_j|$$

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

ORS Research Design & Data Analysis Lab

Office of Research and Scholarship

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

STARs Award

The University of Texas System
P. Indic (PI, UT Tyler)

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)

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)

Pre-Vent

National Institute Of Health Grant
P. Indic (Analytical Core PI, UT-Tyler)
N. Ambal (PI, Univ. of Alabama, Birmingham)

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

QUESTIONS