A Gentle Introduction To Machine Learning

Kyle Kastner
Southwest Research Institute (SwRI)
University of Texas - San Antonio (UTSA)

Outline

- Motivations
- Broad Categories
- Basic Techniques
- Regression
- Classification
- Resources
- Final Comments

Motivations

- Dense working vocabulary
  - Statistics
  - Mathematics
  - Computer Science
  - Domain expertise (biology, audio, etc.)
- "Black box" treatment
- Data, data, data!

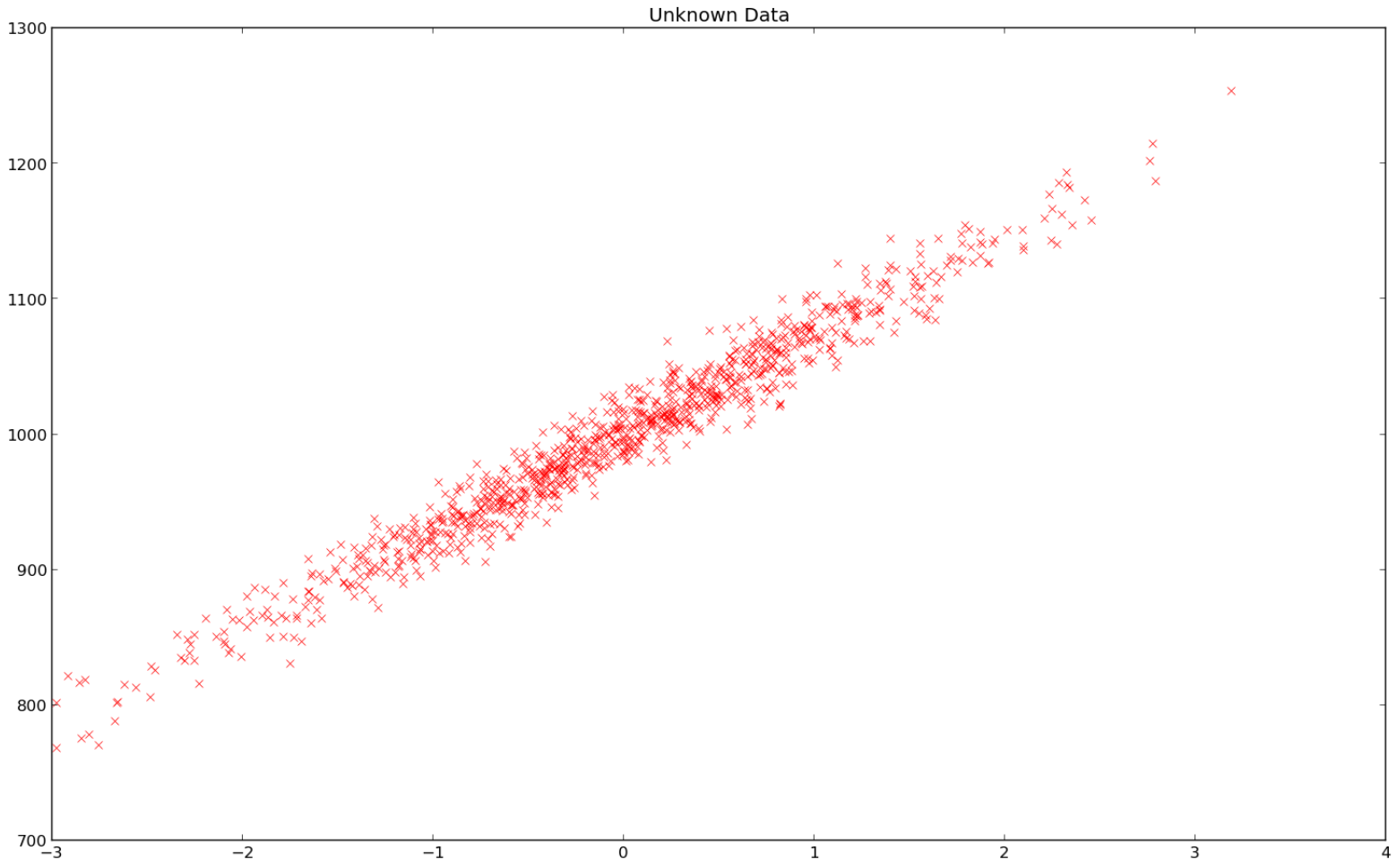
Broad Categories

- Unsupervised (unlabeled training)
  - Data is abundant
  - Proper preprocessing is complicated
- Supervised (labeled training)
  - Hard to find domain data
  - Harder to gather validated data

Basic Techniques

- Regression
  - Find mathematical generating function
- Classification
  - Differentiate between categorical labels
- Clustering
  - Group similar data (usually unlabeled)
  - Not covered here

Regression



Regression

- Using linear regression
- Find the "best fit" line or polynomial
- Simple implementation
- Outliers will greatly affect results

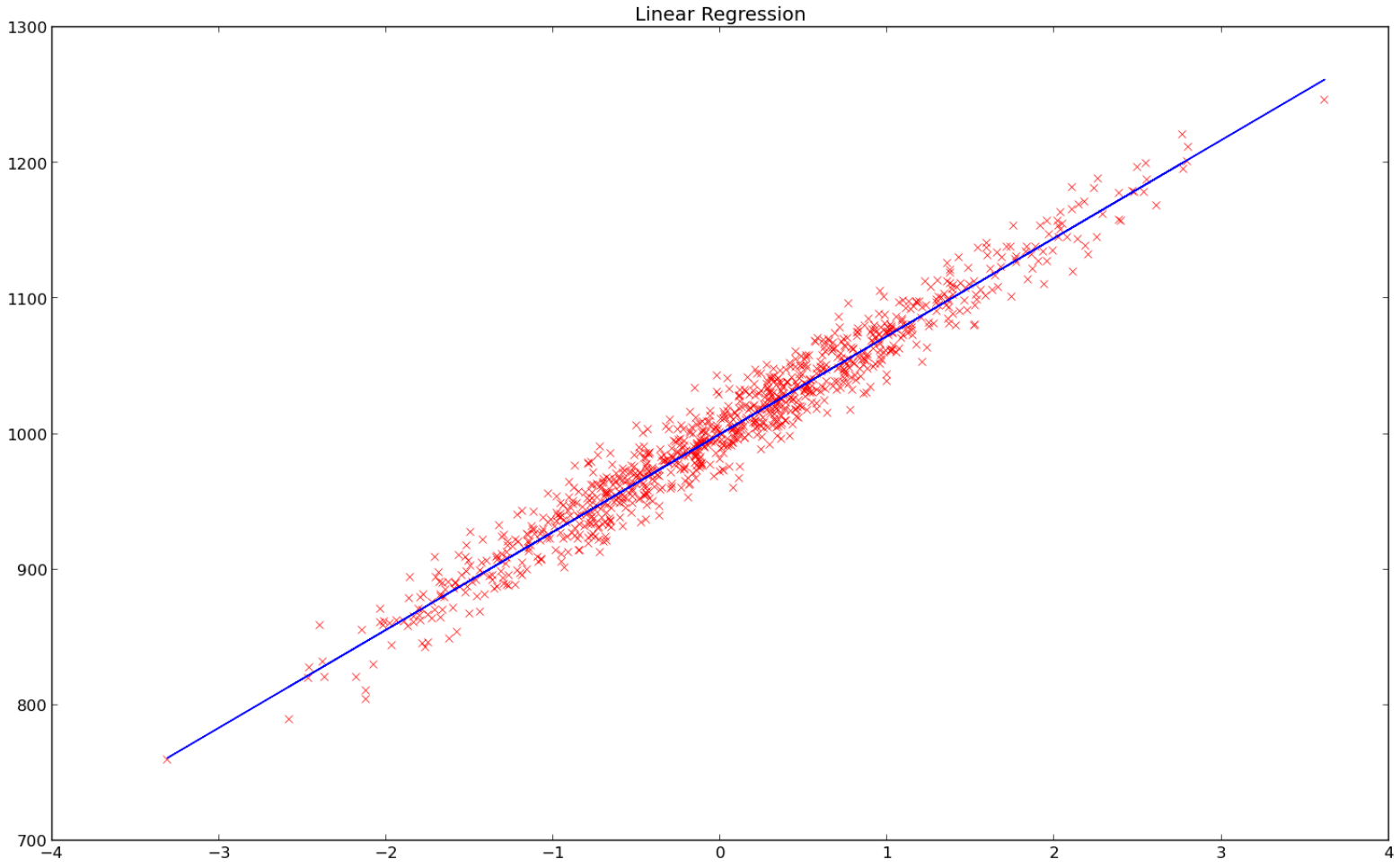
Optimization

- Gradient descent
  - Cost function: $J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$
  - Gradient: $\frac{\partial}{\partial j}J(\theta) = x_j(h_\theta(x) - y)$
  - Update: $\theta_j := \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j$
  - Theta is the parameter to optimize
  - Learning rate (alpha) is called a hyperparameter
- scipy.opt, mystic or cvxopt

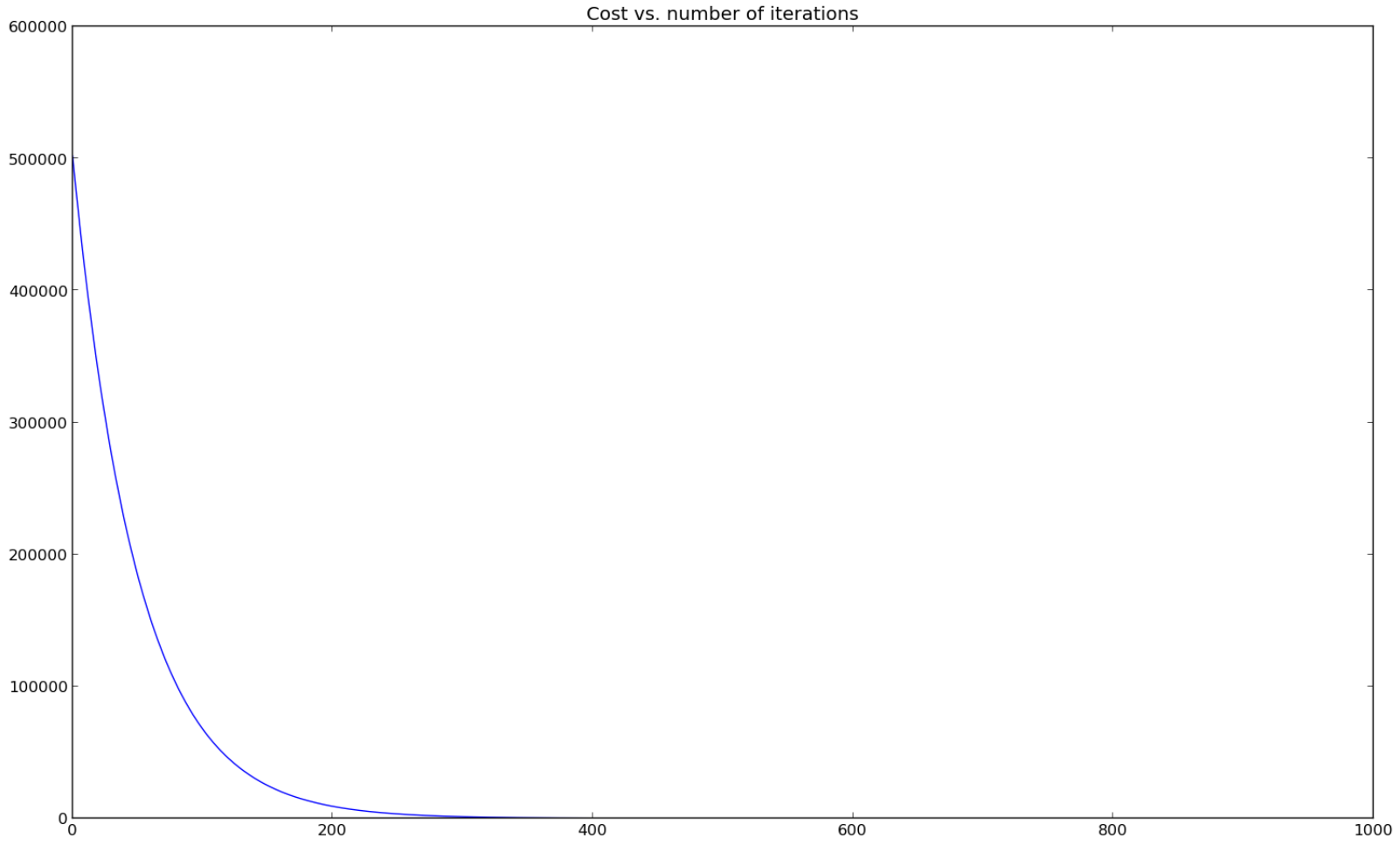
Gradient Descent

- Batch gradient descent
  - Runs over every sample, then updates
- Stochastic gradient descent
  - Sample by sample updates, slower to converge
- Mini-batch gradient descent
  - Adjustable update cost and speed

Regression


Linear Regression
1300
1200
1100
1000
900
800
700
−4
−3
−2
−1
0
1
2
3
4

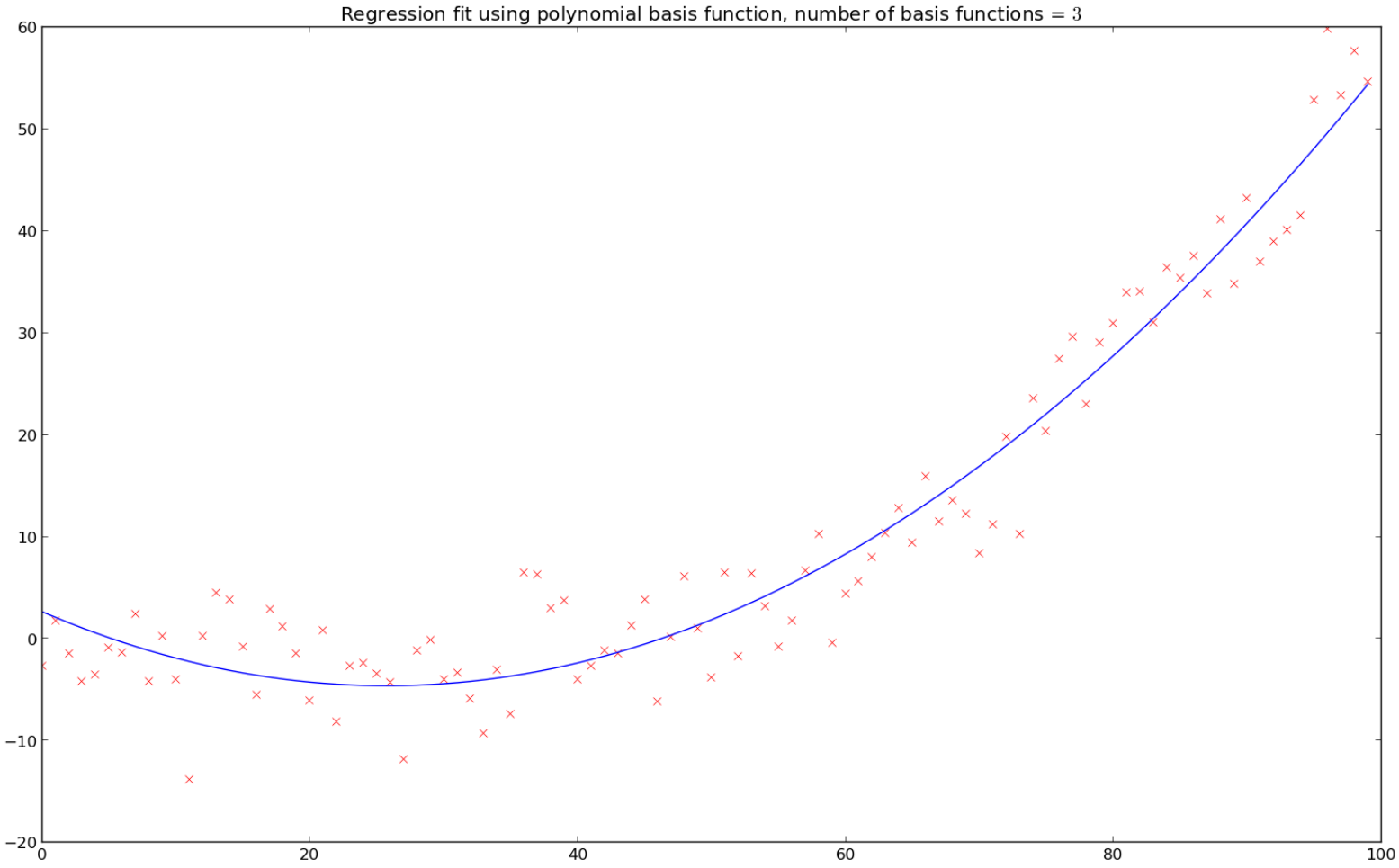
Regression


Cost vs. number of iterations
600000
500000
400000
300000
200000
100000
0
0
200
400
600
800
1000

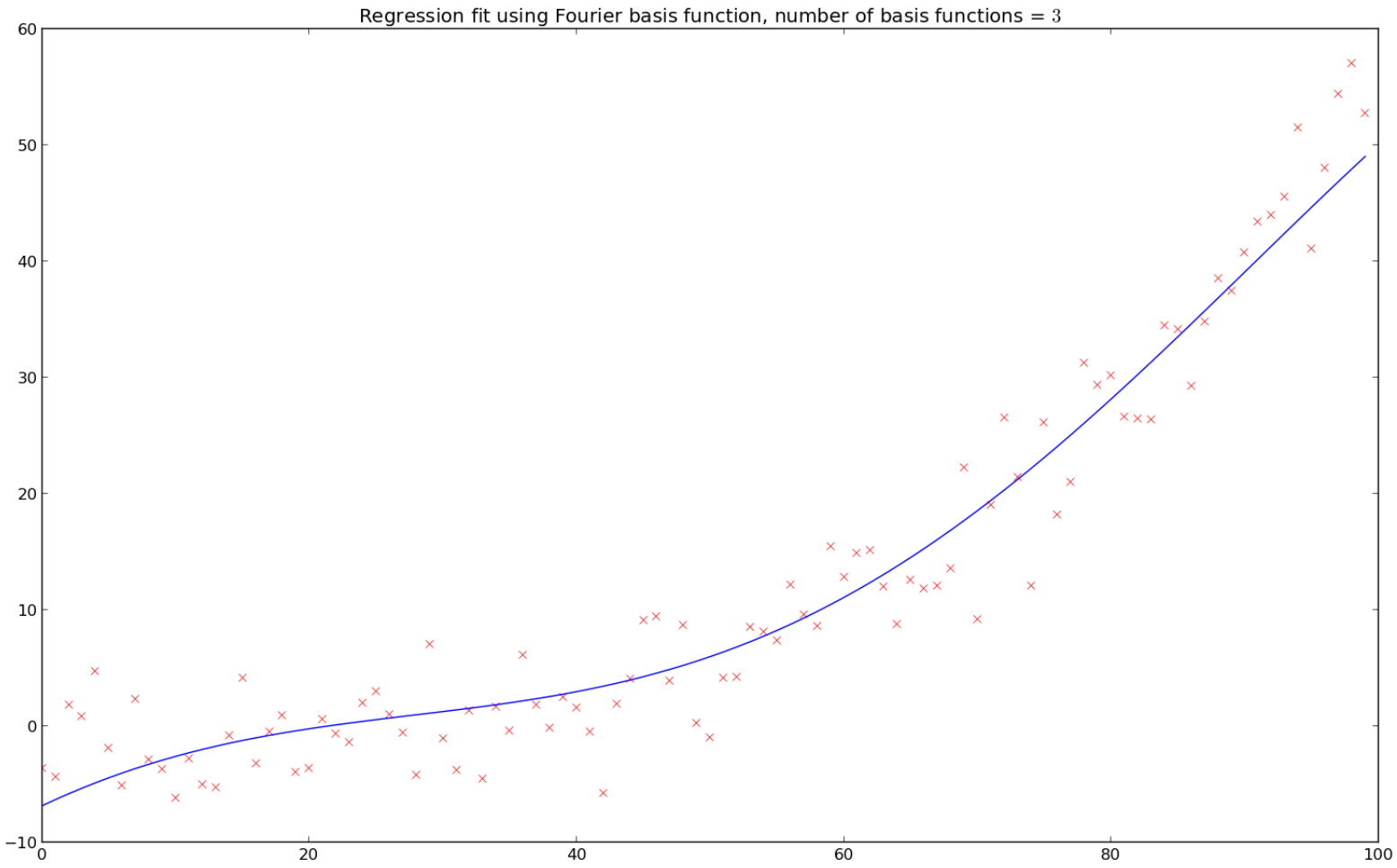
Regression

- Linear regression into different basis
- Basis functions
  - Polynomial
  - Fourier
  - Other
- Alternate optimization method
  - Normal equation
  - Moore-Penrose pseudoinverse

Regression


Regression fit using polynomial basis function, number of basis functions = 3
60
50
40
30
20
10
0
−10
−20
0
20
40
60
80
100

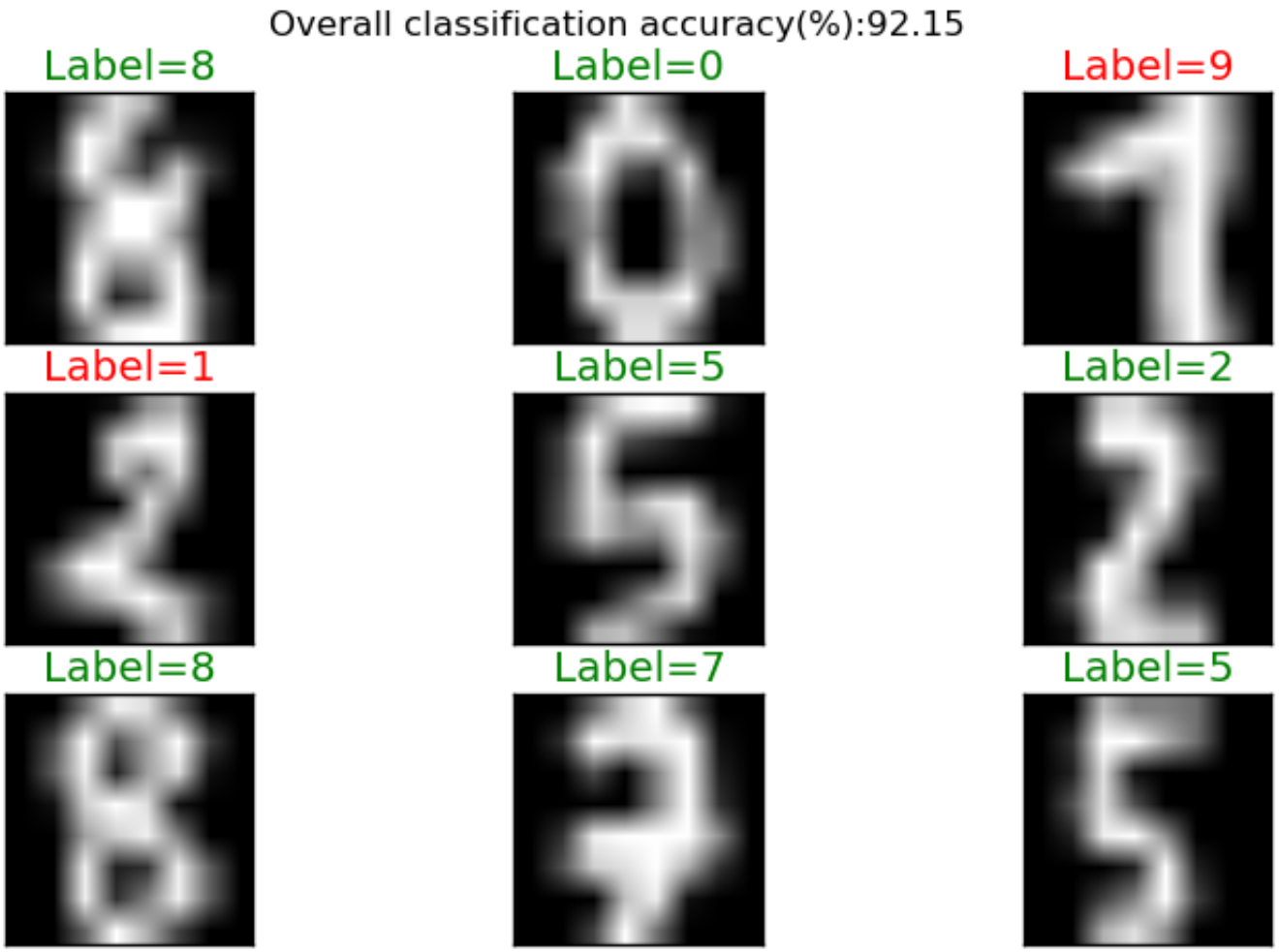
Regression


Regression fit using Fourier basis function, number of basis functions = 3
60
50
40
30
20
10
0
−10
0
20
40
60
80
100

Classification

- Using logistic regression
- Optimize parameters for each class label
  - fmin_cg is conjugate gradient
  - One vs. all, choose most probable (index of max)
- Example with MNIST digits
- 1797 examples of 8x8 pixel digits

Classification



Resources

- Coursera courses
  - Andrew Ng's Machine Learning
  - Geoff Hinton's Neural Networks
- Scikit-learn Documentation and Examples
- Pattern Recognition and Machine Learning
  - Christopher M. Bishop

Final Comments

- Data preprocessing can be important
- Prefer simple models to complex
- Separate training and testing data
- DIY, abandon it on GitHub, then use scikit-learn

Questions?

Code on GitHub:
https://github.com/kastnerkyle/SciPy2013