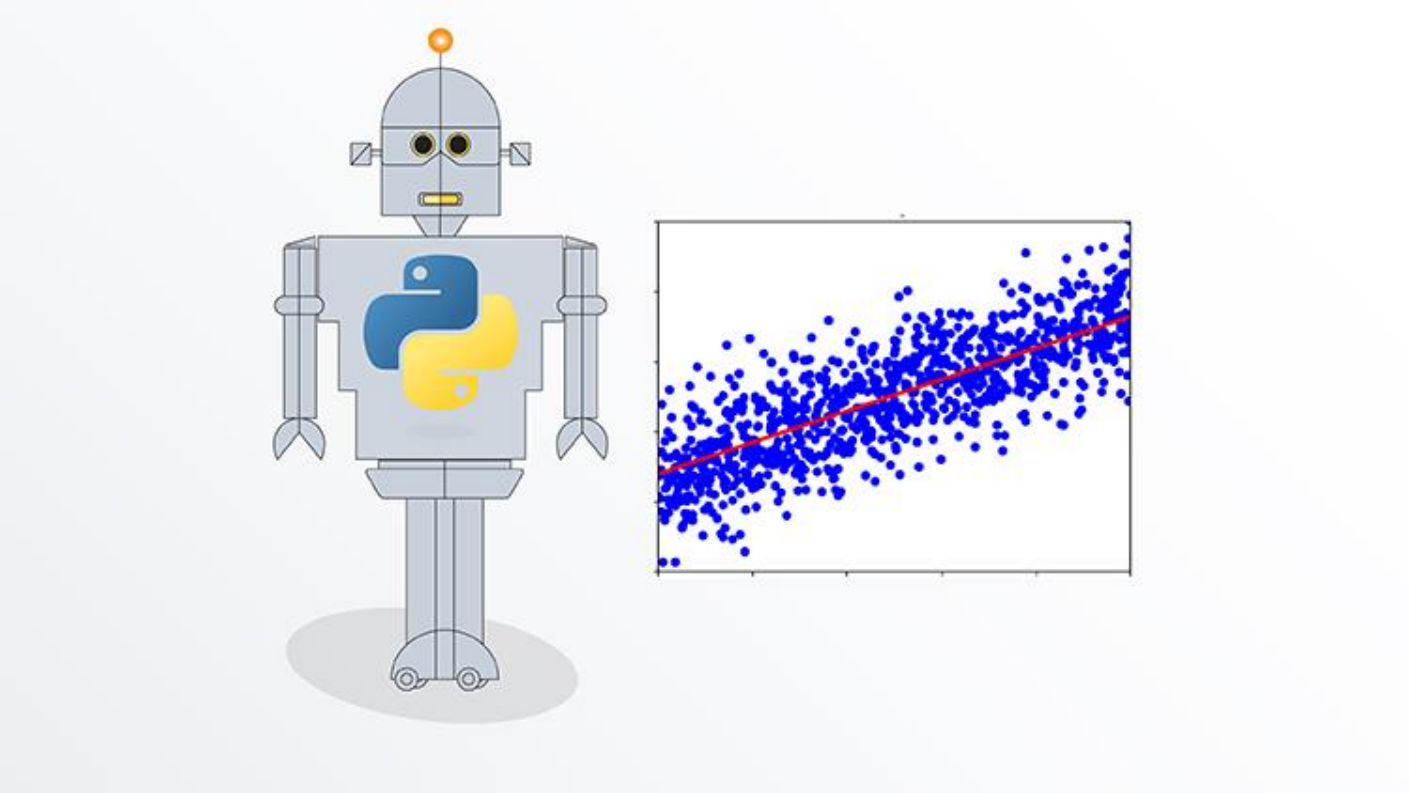
What is Machine Learning?

Chapter 2: Supervised Machine Learning



Supervised Machine Learning

Definition (adapted from Wikipedia page on Supervised Learning)

- Supervised Learning is task of learning a function that maps an input to an output based on example input-output pairs. The learned function is used to predict outputs for new input values

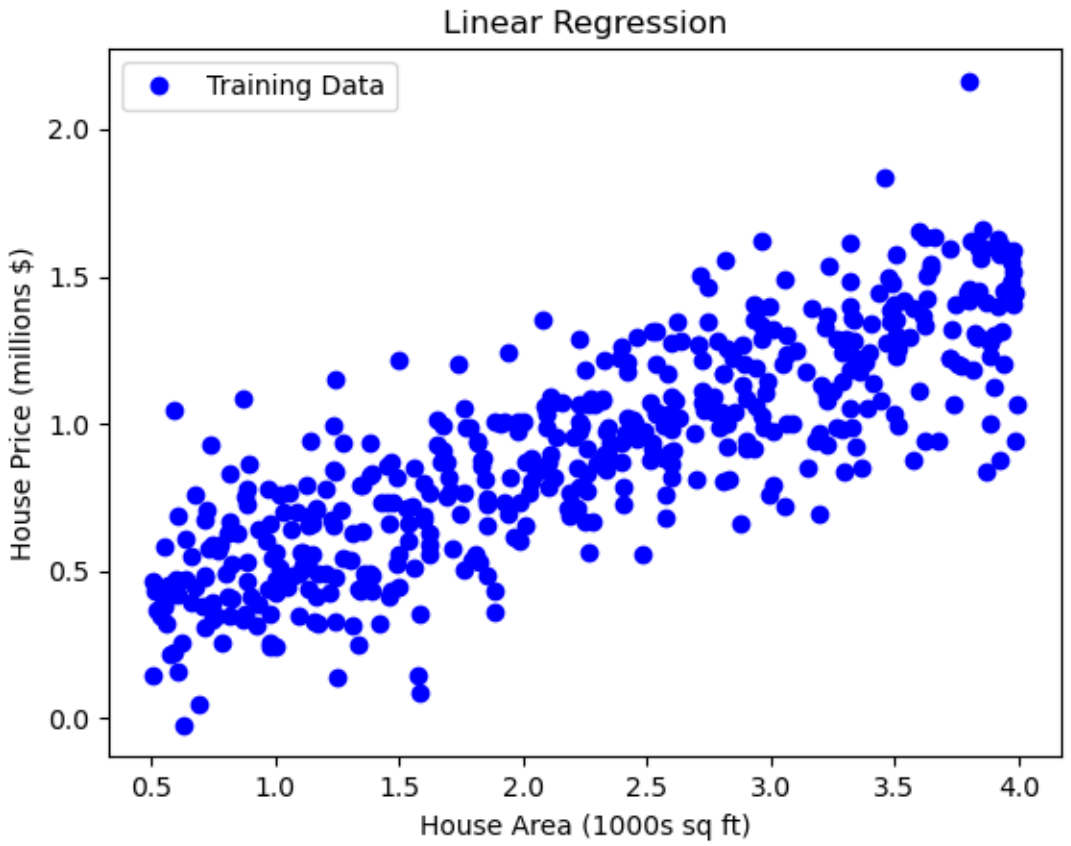
Informal Definition:

- Learn a function that maps inputs to outputs based on given data
- Use function to make predictions

House Price Prediction Using Linear Regression

Problem:

- Find straight line to fit house price data (price versus area)



Linear Regression: Solution Approach

(1) Data:

- For each data point: input X is house area, output Y is house price

(2) Function mapping inputs to outputs:

- Assume linear function: Y = W*X + b and find best slope W, intercept b

(3) Loss Function:

- Loss measures “goodness of fit” of linear function by comparing actual Y values to those predicted by linear function
- Typically use mean squared error function for linear regression

(4) Learning (also called training or fitting in ML literature):

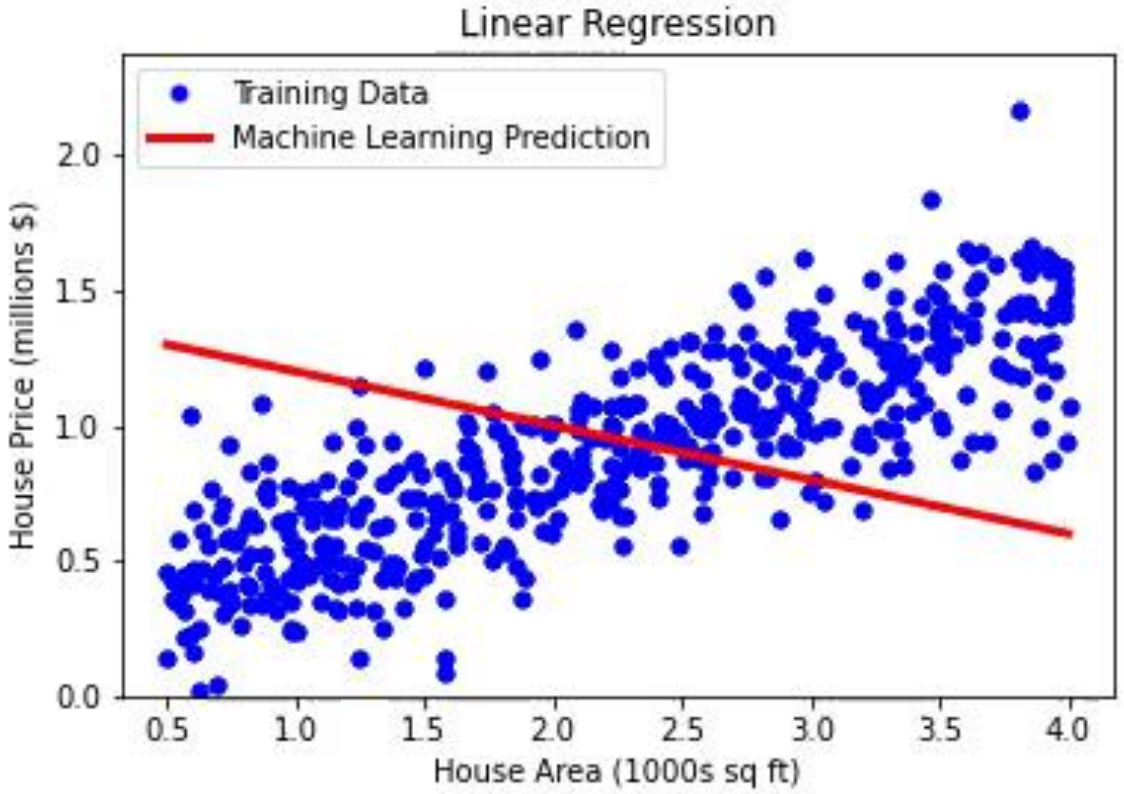
- Use optimization algorithm to find W, b that minimize loss function for training data

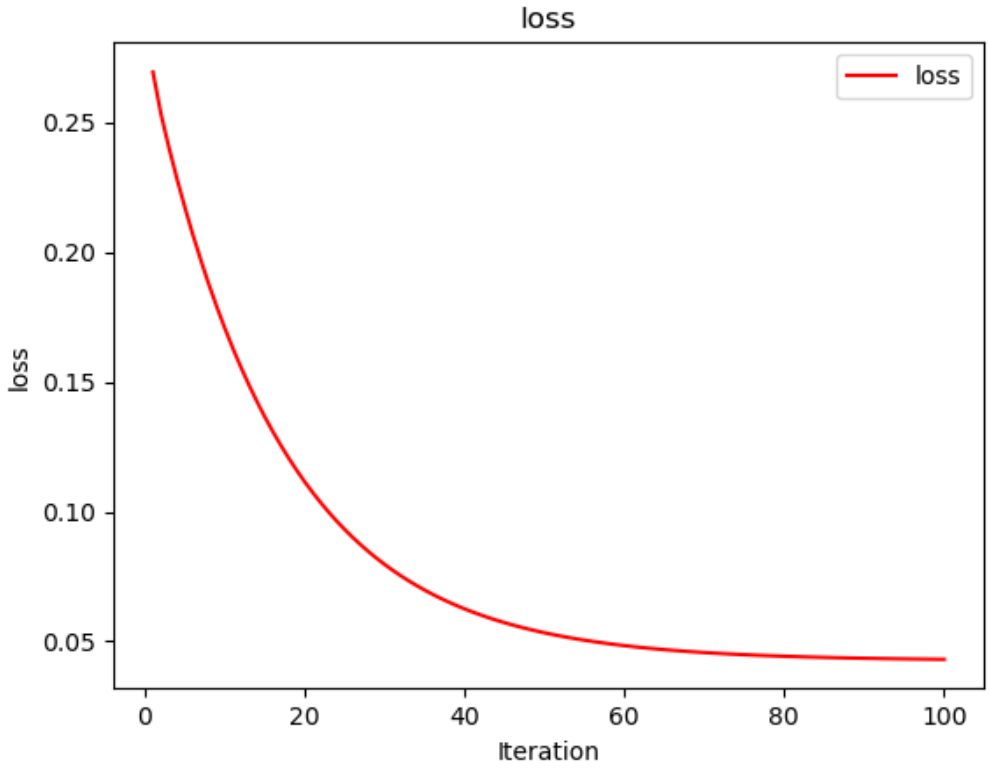
(5) Prediction:

- Use learned W, b and function in (2) to predict house prices if new house area data is provided

Linear Regression: Results

- House price data and machine learning prediction
- Loss function over 100 iterations





Supervised Learning: General Approach

(1) Data:

- Typically input is vector X (d entries/features), output is Y (single value)

(2) Function:

- Assume general form of function for mapping training data input to output
- General form has many parameters (analogous to slope and intercept)

(3) Loss Function:

- Measure goodness of fit of function in (2) using a loss function

(4) Learning (Training):

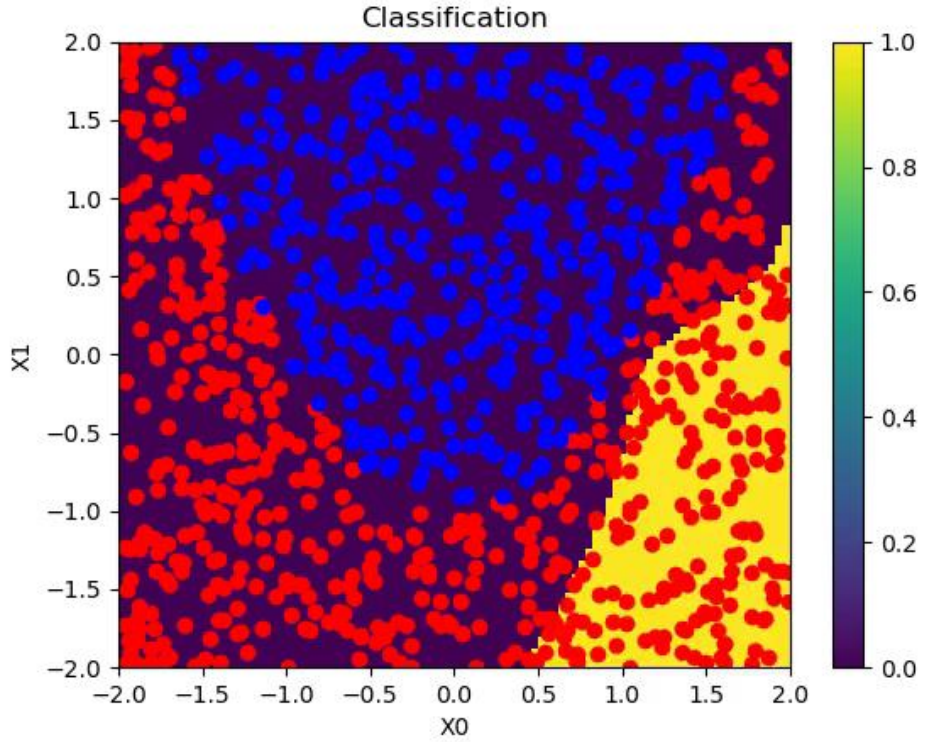
- Use optimization algorithm to find parameter values that minimize loss function for training data (iterative approach: make initial guess and then improve)

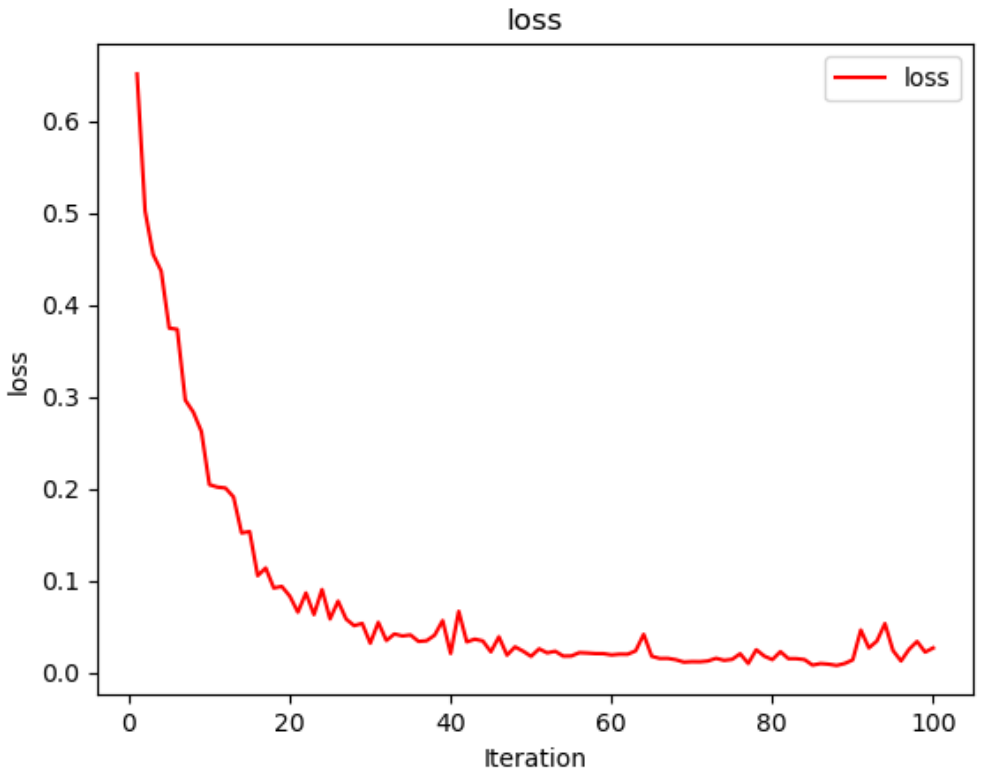
(5) Prediction:

- Use learned parameters and function in (2) to predict output for new X values

Binary Classification

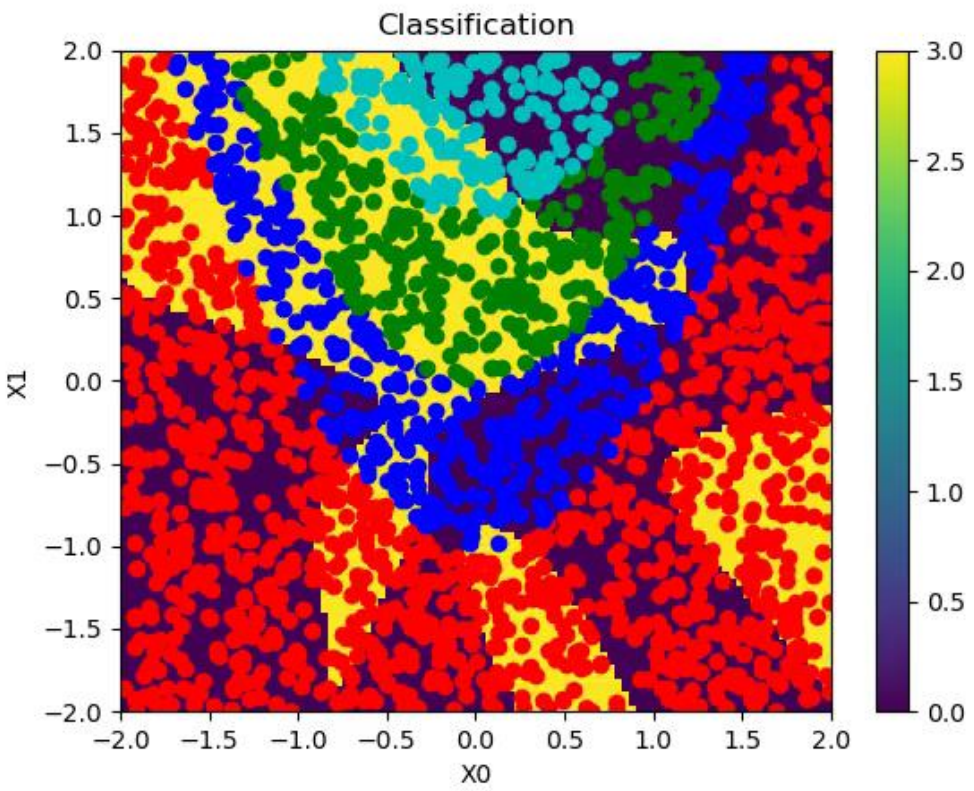
- Data: INPUT: 2 features (x0,x1) coordinates: OUTPUT: label 0 or 1
- Goal: find function that best fits data
- Loss Function: shows progression of learning

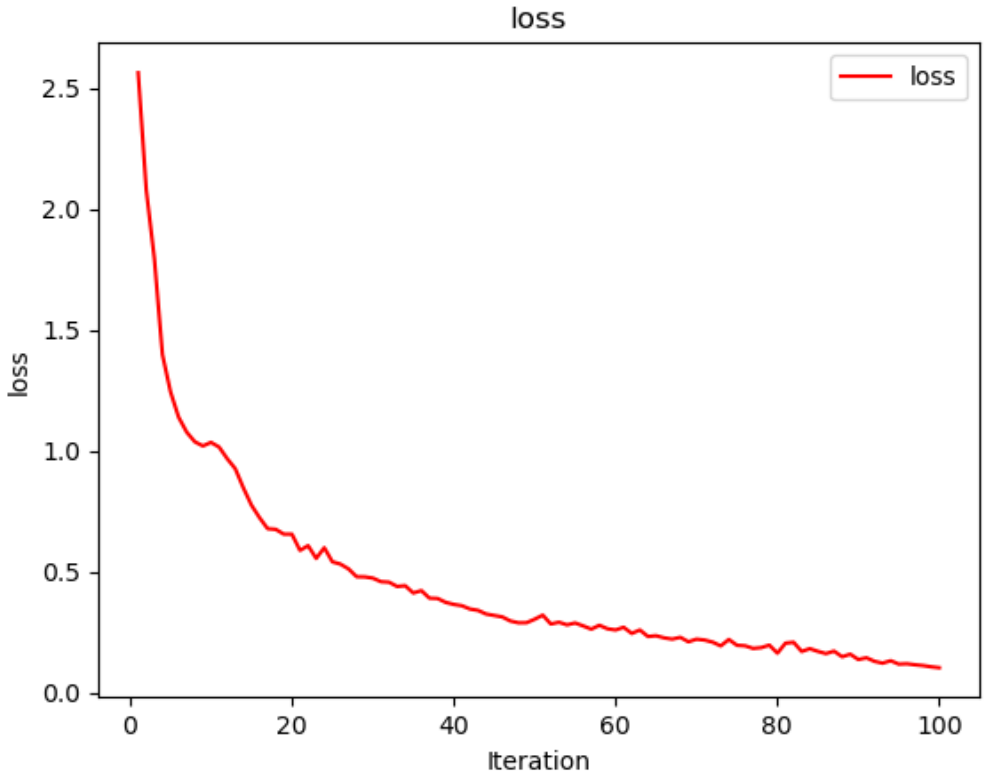




Multi-Class Classification

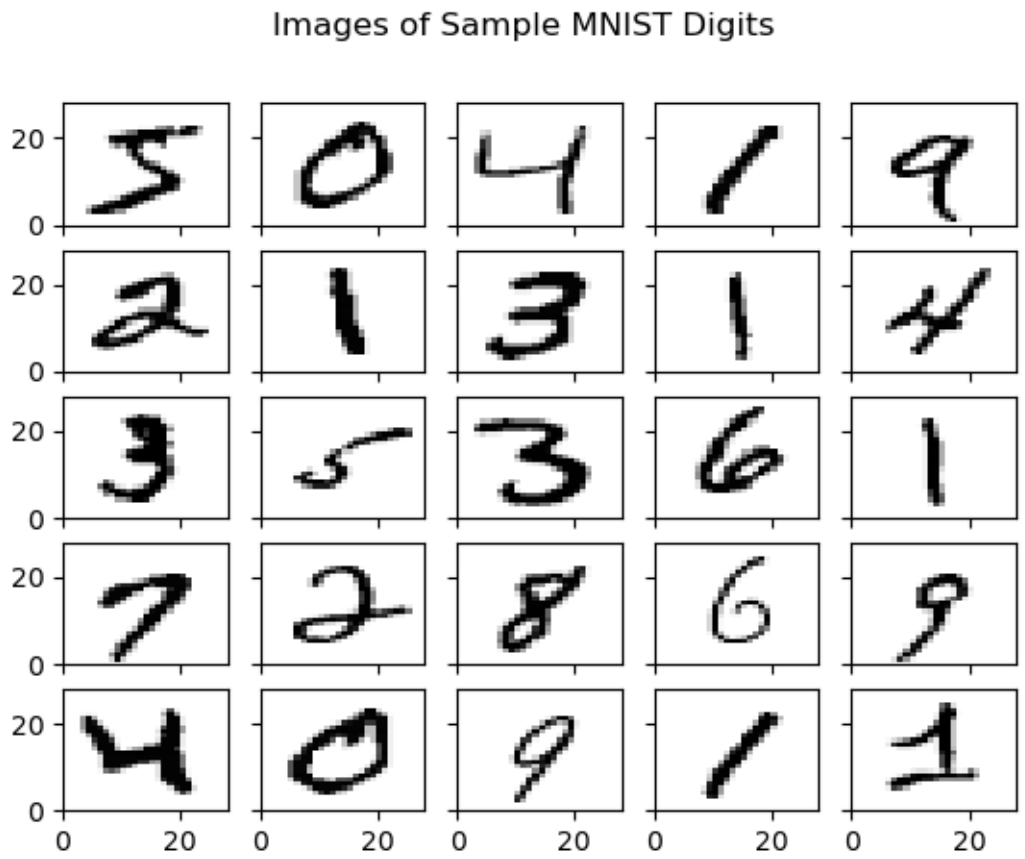
- Data: INPUT: 2 features (x0,x1) coordinates: OUTPUT: label 0, 1, 2, or 3
- Goal: find function that best fits data
- Loss Function: shows progression of learning





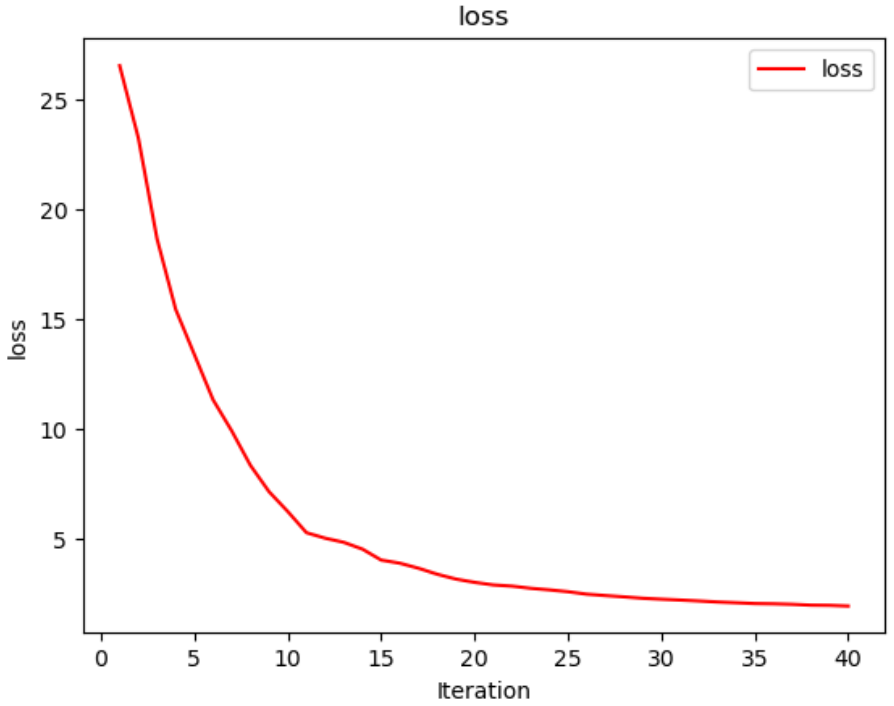
MNIST Digits Classification

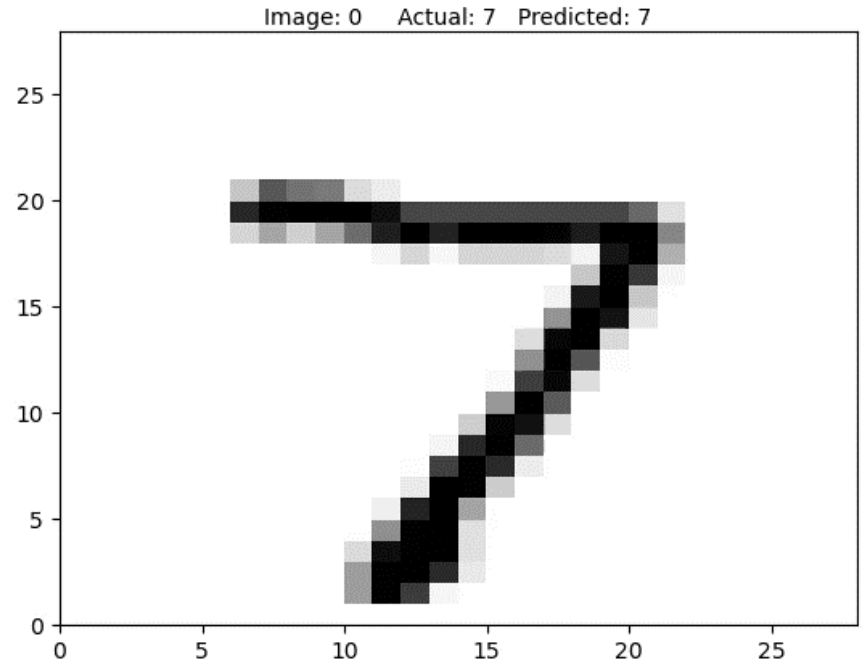
- Thousands of handwritten digit images with 28x28 resolution
- Data Source: http://yann.lecun.com/exdb/mnist/



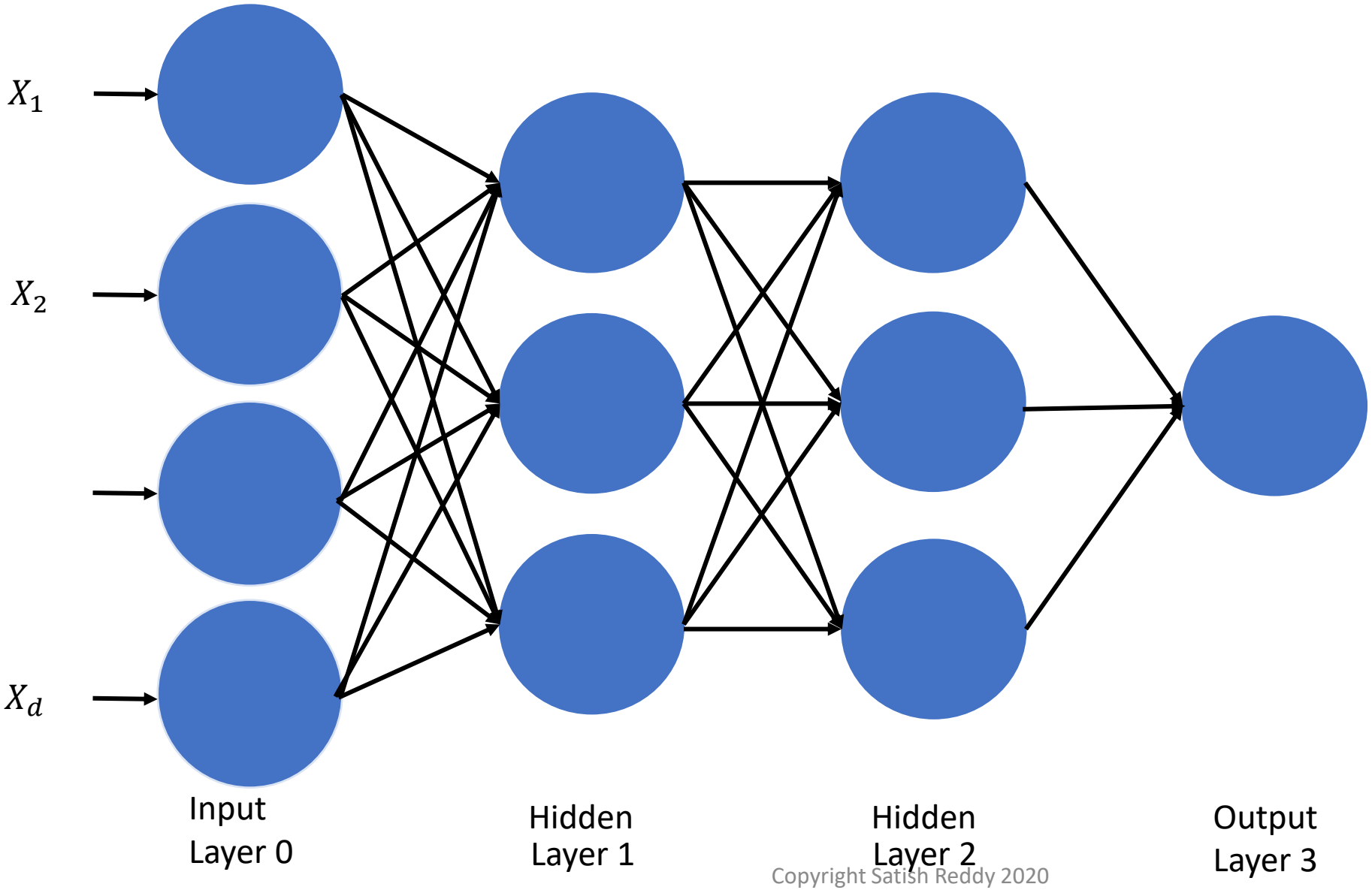
MNIST Digits Classification Results

- X for each image is vector of 784=28x28 pixel intensities, Y is the digit 0,1,2,…, or 9
- Use 6000 images in data set
- Function mapping inputs to outputs fit to data has more than 100,000 parameters
- Movie shows Actual and Predicted Digit Labels





Neural Network (3 Layer)



- Neural Network (NN) is way for defining complex functions to map inputs to outputs
- X is input at layer 0 and estimate for Y is output at final layer
- Layers are connected by matrix multiplications and activation functions
- For Regression/Binary Classification output layer has 1 unit
- For Multi-class Classification output layer has multiple units

Supervised Learning: Applications

Application	Type	Input/Output Data
House Price Prediction	Regression	Input: house features (lot size, house area, # of rooms, size of garage, etc) Output: house price
Spam Filter	Binary Classification	Input: text/email messages Output: label spam or not spam
Image Classification	Binary or Multi-Class Classification	Input: images (pixel intensities) Output: label for each image
Language Translation	Multi-Class Classification	Input: English words Output: French translations

Supervised Learning: Notes

Component	Notes:
Data	Natural Language Processing: • Various approaches for converting words/sentences/documents into numbers Image Classification: • Use pixel intensities and image augmentation to expand data set
Function	Functions can be built using Neural Networks Natural Language Processing: • Use Recurrent Neural Networks or other specialized structures Image Classification: • Use Convolutional Neural Networks or other specialized structures Alternatives to Neural Network: Decision Trees, Random Forest, Boost
Loss Function	• Regression Problems: means squared error • Binary Classification: binary cross entropy • Multi-class Classification: cross entropy
Learning	• Approaches for minimizing loss function: Gradient Descent, Adam, and others • May not converge to minimum or may converge to local minimum