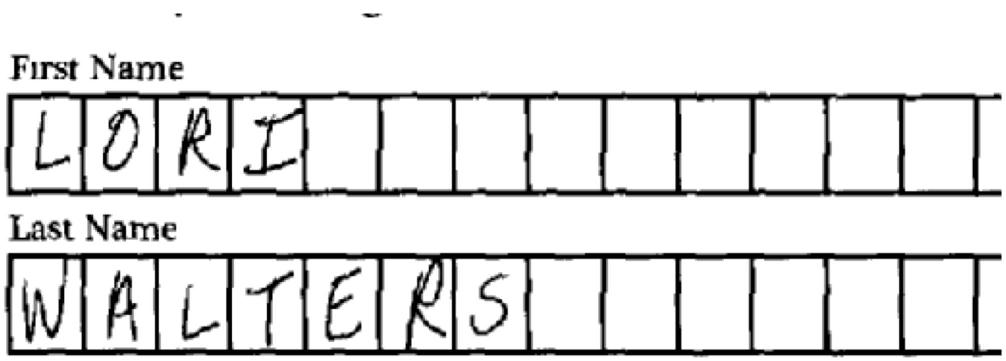
Introduction to Machine Learning

Jay Urbain, PhD

Credits: NYU, nVidia DLI

Machine Learning

- Machine Learning is the ability to teach a computer without explicitly programming it
- Examples are used to train computers to perform tasks that would be difficult to program

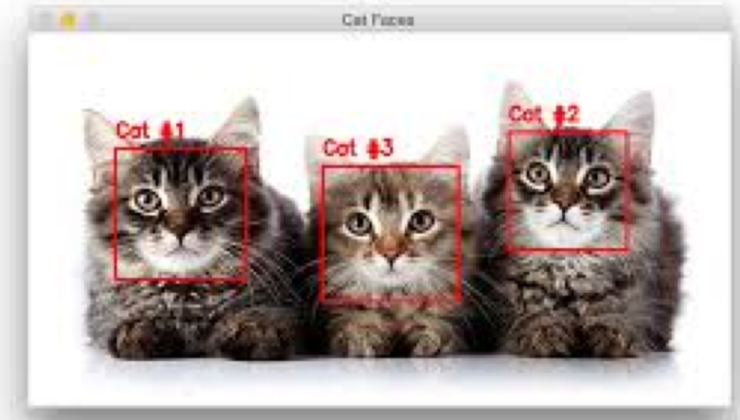


First Name

LORI

Last Name

WALTERS



Types of Machine Learning

- Supervised Learning
    - Training data is labeled
    - Goal is correctly label new data
- Reinforcement Learning
    - Training data is unlabeled
    - System receives feedback for its actions
    - Goal is to perform better actions
- Unsupervised Learning
    - Training data is unlabeled
    - Goal is to categorize the observations

Applications of Machine Learning

- Handwriting Recognition
  - convert written letters into digital letters
- Language Translation
  - translate spoken and or written languages (e.g. Google Translate)
- Speech Recognition
  - convert voice snippets to text (e.g. Siri, Cortana, and Alexa)
- Image Classification
  - label images with appropriate categories (e.g. Google Photos)
- Autonomous Driving
  - enable cars to drive

Features in Machine Learning

- Features are the observations that are used to form predictions
  - For image classification, the pixels are the features
  - For voice recognition, the pitch and volume of the sound samples are the features
  - For autonomous cars, data from the cameras, range sensors, and GPS are features

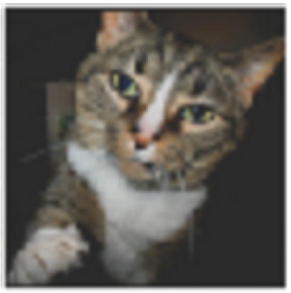
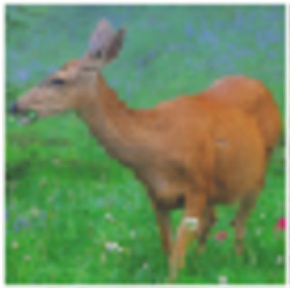
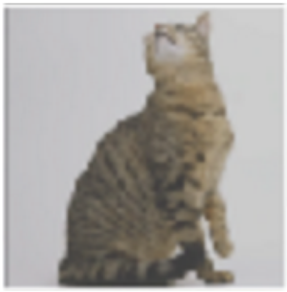
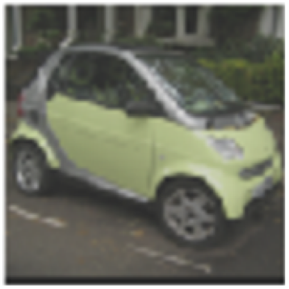
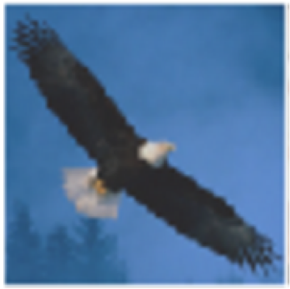
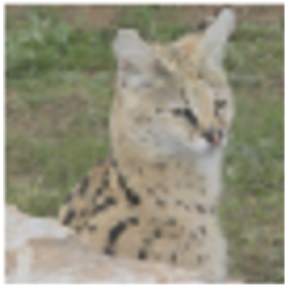
- Extracting relevant features is important for building a model
  - Time of day is an irrelevant feature when classifying images
  - Time of day is relevant when classifying emails because SPAM often occurs at night

- Common Types of Features in Robotics
  - Pixels (RGB data)
  - Depth data (sonar, laser rangefinders)
  - Movement (encoder values)
  - Orientation or Acceleration (Gyroscope, Accelerometer, Compass)

Measuring Success for Classification

- True Positive: Correctly identified as relevant
- True Negative: Correctly identified as not relevant
- False Positive: Incorrectly labeled as relevant
- False Negative: Incorrectly labeled as not relevant

Example: Identify Cats

Prediction:	+	−	−	+	−	+
Image:						
	True Positive	**True** Negative	False Negative	False Positive		

Images from the STL-10 dataset

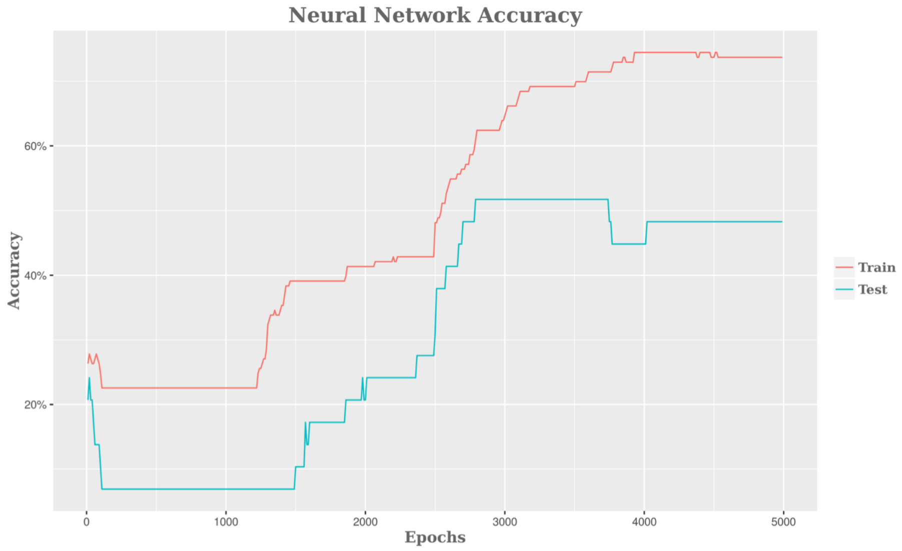
Precision, Recall, and Accuracy

- Precision
  - Percentage of positive labels that are correct
  - Precision = (# true positives) / (# true positives + # false positives)
- Recall
  - Percentage of positive examples that are correctly labeled
  - Recall = (# true positives) / (# true positives + # false negatives)
- Accuracy
  - Percentage of correct labels
  - Accuracy = (# true positives + # true negatives) / (# of samples)

Training and Test Data

- Training Data
  - data used to learn a model
- Test Data
  - data used to assess the accuracy of model

- Overfitting
  - Model performs well on training data but poorly on test data



Bias and Variance

- Bias: expected difference between model’s prediction and truth
- Variance: how much the model differs among training sets

- Model Scenarios
  - High Bias: Model makes inaccurate predictions on training data
  - High Variance: Model does not generalize to new datasets
  - Low Bias: Model makes accurate predictions on training data
  - Low Variance: Model generalizes to new datasets

Supervised Learning Algorithms

- Linear Regression
- Decision Trees
- Support Vector Machines
- K-Nearest Neighbor
- Neural Networks

Supervised Learning Frameworks

Tool	Uses	Language
Scikit-Learn	Classification, Regression, Clustering	Python
Spark MLlib	Classification, Regression, Clustering	Scala, R, Java
Weka	Classification, Regression, Clustering	Java
Caffe	Neural Networks	C++, Python
TensorFlow	Neural Networks	Python