Introduction to Machine Learning

Day 0

Welcome to Machine Learning! Today you'll get yourself introduced with the idea of making a computer learn by itself without being explicitly programmed. You'll also see some classifications of Machine Learning problem, mainly:

- Supervised Learning
  - Linear Regression
  - Logistic Regression
  - Linear Discriminant Analysis
  - K-Nearest Neighbors
  - Decision Trees
- Unsupervised Learning
  - K-Means Clustering
  - Principal Component Analysis
  - Anomaly Detection
  - Collaborative filtering and Recommender Systems
- Reinfocement Learning

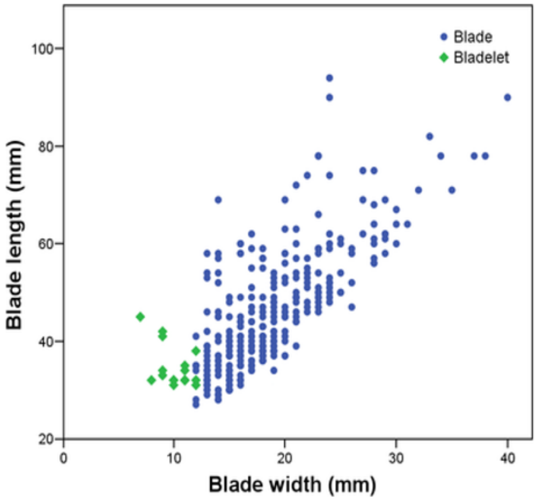
You can find the relevant materials for today in the links below:

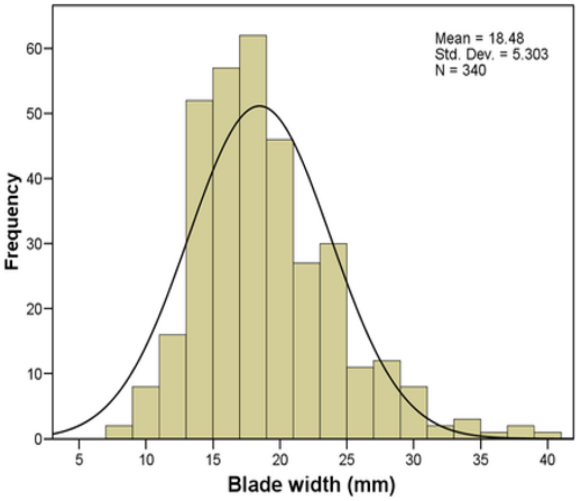
- What Is Machine Learning
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Day 1

This day, you'll get to know the most widely used tools for Machine Learning that will speed up your Machine Learning algorithms with the help of cleverly optimized Linear Algebra libraries. This is possible because of a process called "Vectorization" which exploits the benefits of SIMD (Single Intruction Multiple Data) Parallelism, which speeds up your algorithms by a factor of 400-600. Mainly, the tools that you'll be using through this course are:

- Python
  There are some linear algebra libraries of python that you'll get accustomed to throughout this course.
  - SciPy
  - NumPy
  - Matplotlib
- JupyterLab
- Conda
- Examples of Data Visualization using Matplotlib
- Basic Linear Algebra for Machine Learning



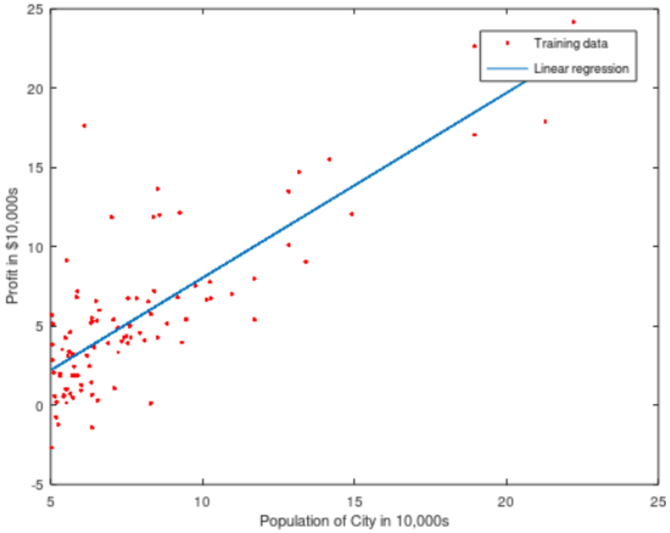


There are some links that you will need to visit frequently while exercising ML algorithms, and these are listed below for your convenience:

- Numpy Documentation
- Matplotlib Documentation
- Python Documentation
- JupyterLab Documentation
- Conda Introduction

Day 2

Today you will learn basic Linear Regression and Gradient Descent, as well as some optimization for your learning algorithms including Feature Scaling and Mean Normalization. You'll need to have a brief idea of basic differential calculus for Gradient Descent as a pre-requisite. Using Linear Regression and Gradient Descent, you can fit a linear model on a sample dataset as follows:

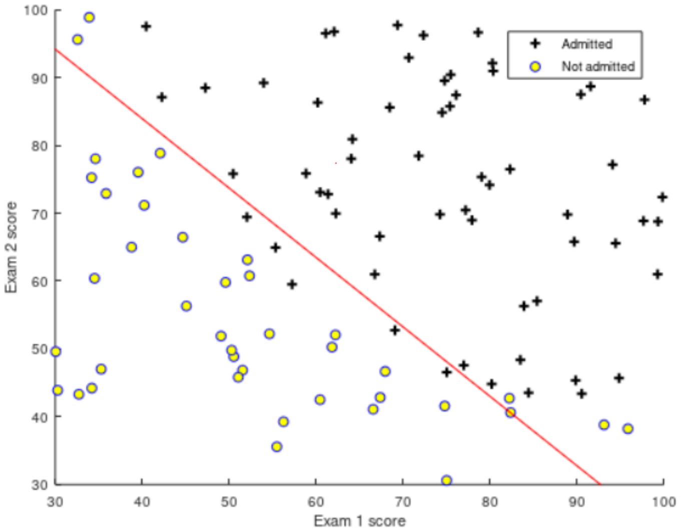


List of relevant information and videos:

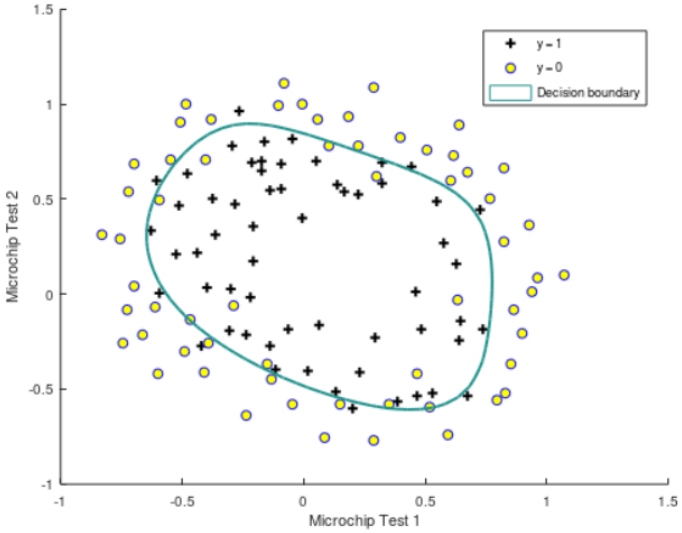
- Representing the model
- Squared error cost function
- Feature scaling and Mean normalization
- Gradient Descent

Day 3

Logistic regression is a classification algorithm. Unlike regression algorithms which gives a continuous valued output, classification algorithms provide discrete valued output which can be used for creating decision boundaries and digit recognition.



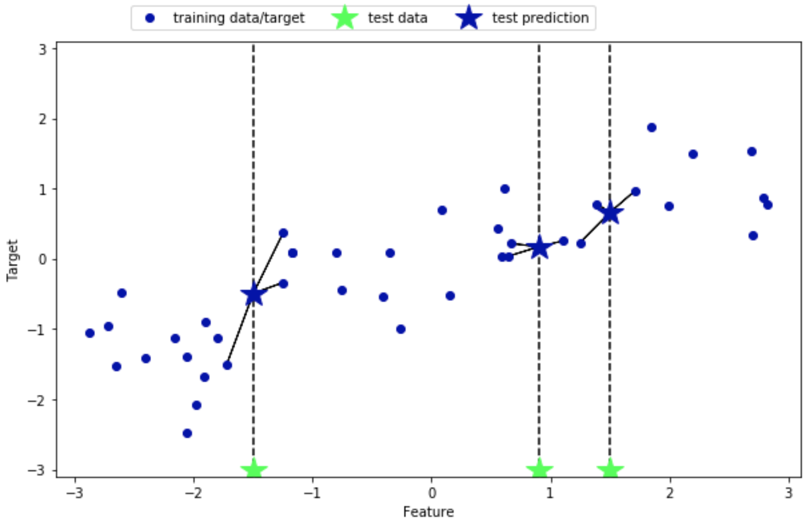
Further optimizing the learning algorithms by using multiple features can be used to find out non-linear decision boundaries.



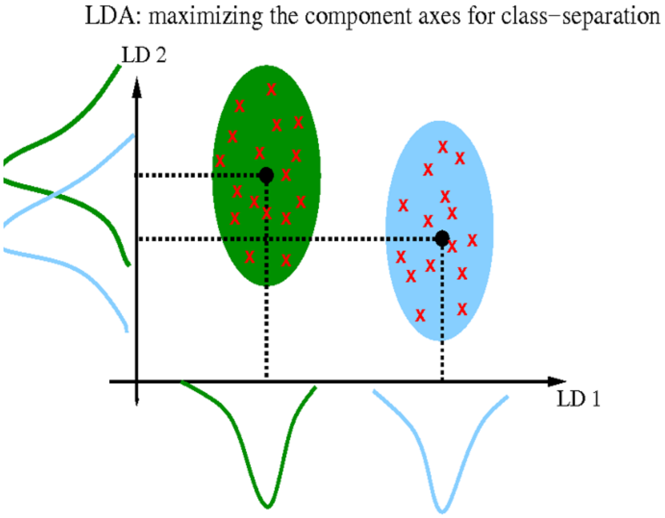
- Logistic Regression concept
- Decision boundary
- Cost function for Logistic regression
- Bias and Variance
- Regularization

Day 4

More supervised learning algorithms will give you a brief overview of Linear Discriminant Analysis, K-Nearest Neighbors and Decision trees.



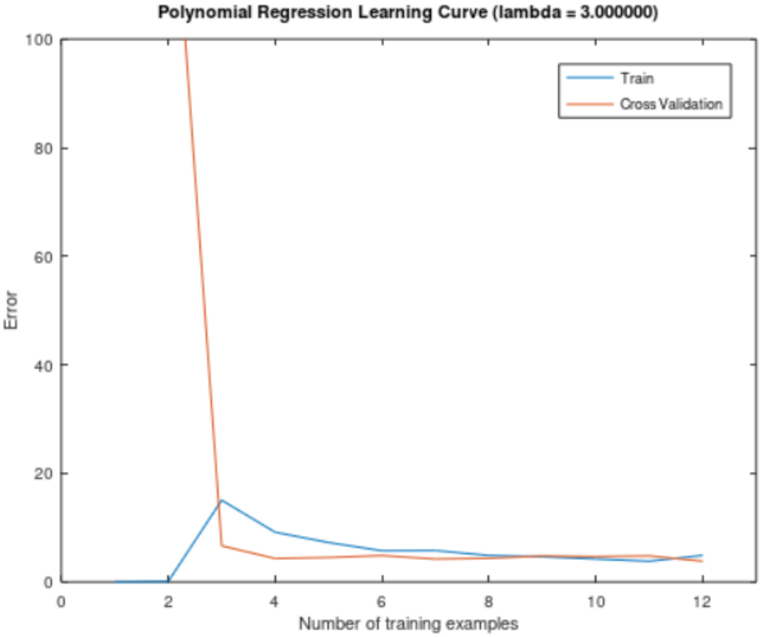
K-Nearest Neighbors



- K-Nearest Neighbors
- Linear Discriminant Analysis
- Decision Trees

Day 5

Measuring the performance of a learning algorithm is undoubtedly the most crucial part of a Machine Learning project. Using Matplotlib, learning curves can be plotted and be checked against overfitting/underfitting problems. Regularization can be addressed to tackle these issues. Exploiting the benefits of real-valued performance measurement and error metrics such as: Precision, Recall and F1 Score, we can further extract out the best result of our Machine Learning project by choosing the best algorithm.


Polynomial Regression Learning Curve (lambda = 3.000000)
Train
Cross Validation
Error
Number of training examples