Machine Learning Lecture-1

Programming Club

Indian Institute of Technology Kanpur

pclubiitk@gmail.com

February 25, 2016

Acknowledgement

This presentation is meant only for teaching purposes as a part of IIT Kanpur club activity. Most informations used here are taken from different sources available online and hence contents of this presentation should NOT be used anywhere without proper consents from the original sources.

What is Machine Learning?

Machine Learning

Machine Learning is the science of getting computers to learn, without being explicitly programmed.

- Search
- Spam Filters
- Recommendations
- Chat-Bots
- Speech recognition
- Face Detection and Recognition
- Robotics
- And a LOT more!

Supervised and Unsupervised Learning

Supervised Learning

A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. The right answers of the training data are given to us.

There are two types of Supervised Learning

- Classification
- Regression

Unsupervised Learning

Unsupervised learning is the machine learning task of inferring a function to describe hidden structure from unlabeled data.

Linear Regression with One Variable

Linear regression with one variable is also known as "Univariate Linear Regression".

Regression

Predict Real-Valued Outputs.

Univariate Linear Regression

Univariate linear regression is used when you want to predict a single output value from a single input value. We're doing supervised learning here, so that means we already have an idea what the input/output cause and effect should be.

Model Representation

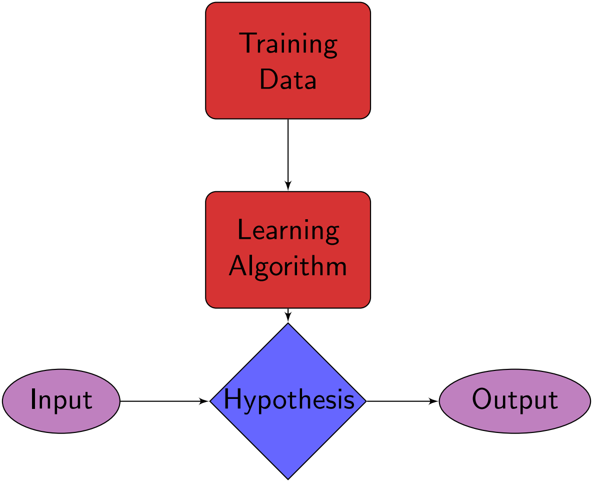
The data is usually categorized into these three parts -

- Training Set
- Validation Set
- Testing Set

Some usual Notations -

- m - Number of Training Examples
- x - Input Features/Variables
- y - Output Features/Variables
- $x^{(i)}$ - ith example of the input
- $y^{(i)}$ - ith example of the output

Hypothesis Function



The Hypothesis Function and Cost Function

The Hypothesis Function

$h_\theta(x) = \theta_0 + \theta_1 x$

We are trying to find a function of this linear from that will reliably map our input data (x) to our output data (y).

The Cost Function

$J(\theta_0, \theta_1) = \frac{1}{2m} * \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

We can measure the accuracy of our hypothesis function by using a cost function. This takes an average of all the results of the hypothesis with inputs from x's compared to the actual output y's.
This function is otherwise called the "Squared error function", or Mean squared error.

Gradient Descent

A way to automatically improve our hypothesis function.

- Consider the graph which has - x- axis as θ_0, z- axis as θ_1 and the vertical y- axis as $J(\theta_0, \theta_1)$.
- From our intuition we know that the most reliable hypothesis would come from the value (θ_0, θ_1) that gives the lowest point of the bottom most pits on our graph.
- We try achieving that graph point by taking the derivative of our cost function.
- The slope of the tangent is the derivative at that point and it will give us a direction to move towards.
- We make steps down that derivative by the parameter α, called the learning rate.

Gradient Descent for Linear Regression

The gradient descent equation

Repeat Until Convergence -
$\theta_j := \theta_j - \alpha * \frac{\partial}{\partial(\theta_j)}(J(\theta_0, \theta_1))$ for j=0 and j=1

This can be reduced to

Gradient Descent for Linear Regression

Repeat Until Convergence -
$\theta_0 := \theta_0 - \alpha * \frac{1}{m}(\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}))$
$\theta_1 := \theta_1 - \alpha * \frac{1}{m}(\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}) * (x^{(i)}))$

Note that we have separated out the two cases for θ_0 and that for θ_1 as we are multiplying $x^{(i)}$ at the end due to the derivative.

Logistic Regression

Logistic Model

To build a classification/Regression model for problems with yes-no (or) negative-positive outcomes.

Logistic Regression Model

$h_\theta(x) = g(\theta^T(x))$
where $g(\tau) = \frac{1}{1+e^{-\tau}}$, (i.e) g is a sigmoid function whose value lies in the range $[0, 1]$

Logistic Regression Model

Decision Boundary

We predict $y = 1$, if $h_\theta(x) \geq 0.5$, this means when $\theta^T x \geq 0$
We predict $y = 0$, if $h_\theta(x) < 0.5$, this means when $\theta^T x < 0$

Consider the case of linear regression model,
where $h\theta(x) = g(\theta_0 + \theta_1(x_1) + .. + \theta_n(x_n))$
To predict whether the output is positive or negative, it is sufficient to check the sign of the argument. This helps in making a very clean analysis.

Logistic Regression - Cost Function

Cost Function

$J(\theta) = \frac{1}{m} * \sum_{i=1}^{m} cost(h_\theta(x^{(i)}), y^{(i)})$
$cost(h_\theta(x^{(i)}), y^{(i)}) = -ylog(h_\theta(x)) - (1-y)log(1-h_\theta(x))$

- The intuition behind this cost function is from the fact that the output is either 1 or 0.
- Moreover, this function is convex in nature thus enabling the gradient descend procedure.

Logistic Regression - Gradient Descend

The gradient descent equation

Repeat Until Convergence -

$\theta_j := \theta_j - \alpha * \frac{1}{m}(\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) * (x_j^{(i)}))$

- The gradient descend looks very similar to that of Linear Regression.
- But, You should remember the hypothesis here is different.

Some Optimizations that can be done -

- Feature Scaling
- Choice of Learning Rate α

Some Pointers to master ML

- Almost everything you would want to do, will have a package already built by somebody.
- DO NOT start everything from scratch.
- While implementing codes in octave/matlab, loop through only if there is NO other alternative.
- Some useful kits you would like to refer to - scikit, TensorFlow

Question

Try this question for sure. This will help you understand the concepts better. Most importantly, it will help in retaining the stuff we did in a hurry. That will be needed when u get into more tougher concepts.

References

Andrew NG, Stanford University, coursera.org
Machine Learning - Online Course

Wikipedia
For Definitions

The End