Machine Learning

Coursera

May 22, 2016

1 Introduction:

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

Machine Learning:
– Grew out of work in AI
– New capability for computers

Examples: – Database mining:
Large datasets from growth of automaton/web
e.g., Web click data, medical records, biology, engineering
– Applications can't program by hand.
e.g., Autonomous helicopeter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.
– Self-customizing programs
e.g., Amazon, Netflix product recommendations
– Understanding human learning (brain, real AI).

1.1 Supervised Learning

e.g. Housing price prediction

Definition: Supervised Learning: "right answers" were given. **e.g. house prices were** *given* **in the outsetthe** "That is, we gave it a data set of houses in which for every example in this data set, we told it what is the right price so what is the actual price that, that house sold for and the toss of the algorithm was to just produce more of these right answers such as for this new house, you know, that your friend may be trying to sell."

"To define with a bit more terminology this is also called a regression problem and by regression problem I mean we're trying to predict a continuous value output. Namely the price."

Regression: Prediction is trying to predict a CONTINUOUS VALUED OUTPUT e.g. price

e.g. Is breast cancer malignant or benign??
Tumor size vs. Malignant, and being binary.

This is a CLASSIFICATION problem. e.g. 0 or 1; malignant or benign; DISCRETE VALUE OUTPUT....

"Support Vector Machine", deals with an infinity long list of features (of data...).

e.g., Problem 1. "You have a large inventory of identical items. So imagine that you have thousands of copies of some identical items to sell and you want to predict how many of these items you sell within the next three months.(?)"

e.g., Problem 2. "You have lots of users and you want to write software to examine each individual of your customer's accounts, so each one of your customer's accounts; and for each account, decide whether or not the account has been hacked or compromised."

Answer: Treat Problem 1 as a Regression Problem; Problem 2 as a Classification Problem.

1.2   Unsupervised Learning

Supervised Learning, we are told explicitly what is the so-called "right" answer, ("Is it benign or malignant")

UNSupervised Learning, given data that doesn't have any labels or that all has the same label or really no labels. So we're given the data set and we're not told what to do with it and we're not told what each data point is. Instead we're just told, here is a data set. Can you find some structure in the data?

You don't know how the data should be "spilt up".

e.g., Google News (is Unsupervised Learning).

e.g., Understanding genomics
Unsupervised, here's a bunch of data; have "no idea what the answer is", but can you find structure in the data... :-)

e.g.'s Organize computer clusters; Social network Analysis; Market Segmentation; Astro image data (!!!!)

Using Octave.

Octave: single value decomposition; but that turns out to be a linear algebra routine, that is just built into Octave.

"What I've seen after having taught machine learning for almost a decade now, is that, you learn much faster if you use Octave as your programming environment, and if you use Octave as your learning tool and as your prototyping tool, it'll let you learn and prototype learning algorithms much more quickly. "

Review Question:
Of the following examples, which would you address using an UNsupervised Learning Algorithm?? NOT: Spam fliter, NOT diabetes diagnois. YES: News articles on the web; YES: customer data for market segments.

2 Linear Regression with One Variable

"Linear regression predicts a real-valued output based on an input value. We discuss the application of linear regression to housing price prediction, present the notion of a cost function, and introduce the gradient descent method for learning."

2.1 Model and Cost Function: Model Representation

e.g. Housing Pricing; Supervised Learning, Regression problem.

$m =$ number of training examples
$x =$ "input" variables/features
$y =$ "output" variables/"target" variable
(x, y) – one training example
$(x^{(i)}, y^{(i)})$ – i^{th} training example

"So here's how this supervised learning algorithm works. We saw that with the training set like our training set of housing prices and we feed that to our learning algorithm. Is the job of a learning algorithm to then output a function which by convention is usually denoted lowercase h and h stands for hypothesis. And what the job of the hypothesis is, is, is a function that takes as input the size of a house like maybe the size of the new house your friend's trying to sell so it takes in the value of x and it tries to output the estimated value of y for the corresponding house."

So h is a function that maps from x's to y's.
(h is the "hypothesis", which isn't the best name!!)

How do we represent h??
$h_\theta(x) = \theta_0 + \theta_1 x$
just a simple linear function!!
Linear regression with one variable, x (aka "univariate linear regression" !!).

2.2 Model and Cost Function: Cost Function

θ_0 and θ_1 are the parameter of the model.
Linear regression. Have a training dataset.

Chose θ_0 and θ_1 so that $h_\theta(x)$ is close to y for our training examples (x, y).

So, an e.g., of a cost function:
$J(\theta_0, \theta_1)$
this Cost Function, is the Squared Error Function.

And we want to MINIMIZE IT.

2.3 Model and Cost Function: Cost Function – Intuition 1

Hypothesis:

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

Parameters:

$$\theta_0 \text{ and } \theta_1$$

Cost Function:

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

Goal: Minimize $J(\theta_0, \theta_1)$

2.4 Model and Cost Function: Cost Function – Intuition 2

Ross et al. (2015)

References

Ross N. P., et al., 2015, MNRAS, 453, 3932