Machine Learning

March 27, 2019

(E)xperience, (T)ask, (P)erformance

1 Supervised Learning

With supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between input and output. Sueprvisde learning probleams are categorized into "regression" and "Classification" problems. Regression - The goal of predicting a continuous valued output. Classification problem - The goal to predict a discrete valued output. In a regression problem, we are trying to predict results within a continuous output, meaning we are trying to map our inpput variables into a continuous function. In a classification problem we are instead trying to predict results in a discrete output.

Example 1:

Given data about houses on the market, try to predict thir price. Price as a function of size is a contionuous output, thus a regression problem.

This could be made into a classification problem if we made the output whether the house sells for more or less than the asking price.

Example 2:

(a) Regression - Given a picture of a person, we have to predict their age on the basis of the given picture

(b) Classification - Given a patient with a tumor, we have to predict whether the tumor is malignant or benign.

2 Unsupervised Learning

Unsupervised Learning deals with a data set with no given information and tries to group it into a pattern. Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.

3 Model Representation

m = Number of training variables. x's = "input" variable / features y's = output variable / target variable (x, y) - one training example (x^i, y^i) i^{th} training example

training set -¿ learning algorithm -¿ h

3.1 Representing the hypothesis(h)

h maps from x's to y's

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

shorthand h(x)

Linear resgression with one variable (univariate linear regression) We will also use X to denote the space of input values, and Y to denote the space of output values. In this example, X = Y = $\mathbb{R}$

4 Cost Function

$\theta_i's$ are parameters. How to choose $\theta_i's$?

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

$\frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^(i) - y^(i))^2)$

Btw... $(\theta_0(x^(i)) = \theta_0 + \theta_1 x^(i))$

In english: You want the sum of i to m (size of training set) of the squared difference beteen prediction of the hypothesis when it is inputed the (house number i) minus the actual price.

"We can measure the accuracy of our hypothesis function by using a cost function. This takes an average difference (actually a fancier version of an average) of all the results of the hypothesis with inputs from x's and the actual output y's."

Basically what we want is the mean of the difference between θ_1 (the slope, so it iterates) and the actual point.

5 Gradient Descent

So we have our hypothesis function and we have a way of measuring how well it fits into the data. Now we need to estimate the parameters in the hypothesis function. That's where gradient descent comes in.

Imagine that we graph our hypothesis function based on its fields θ_0 and θ_1

(actually we are graphing the cost function as a function of the parameter estimates). We are not graphing x and y itself, but the parameter range of our hypothesis function and the cost resulting from selecting a particular set of parameters.

':=' means assignment

Repeat until convergence.

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

α (alpha) is the learning rate, or size of step.

j=0,1 represents the feature index number.

At each iteration j, one should simultaneously update the parameters $\theta_1, \theta_2 ... \theta_n$

For each iteration you update the theta points to get an accurate new derivitive.

With a fixed step size the descent will automaticaly decrease. With an appropriate step size the descent will converge when the gradient descent is 0.

6 From the quiz

If you have a set of x and y's without an explicit $\theta_0 and \theta_1$ then you can infer based on the tendincies in the data set. From the quiz, there was a question asking for the $\theta_0 and \theta_1$. It gave a chart, from the chart you could see that as x increased y increased. For θ_0 or what θ is when x=0, which wasn't defined, you were to infer the answer because all but one of the answers had θ_0 ¿ θ_1. Based off the tendency you would infer that θ_0 would be less than θ_1

7 Matrices

The inverse of a matrix is called a degenerate.

Matrix transpose: A = [[1, 2, 0], [3, 5, 9]

A^T [[1 3], [2 5], [0 9]]