9/14/2016

PTYS595B -- Machine Learning in Planetary Sciences

NEWS OF THE DAY

Machine learning is chaging every day. Google bought deepmind for a lot of money, they had 1 paper and ~15 people working there at the time. Wavenet is text to speech model. It's generally very hard to make machines sound like humans, e.g. human breathe naturally in speech and machines don't. As the speech gets better it sounds more and more human and we become more and more comfortable with it. The neural networks learn the patterns in our voices and apply them to the text to speech.

LAST CLASS RECAP: linear regression

For cars, y is the price, the x row vectors are the conditions of each car, and beta (b) is the vector of weights applied to each condition.

We used a simple linear model to fit the data ponts -- see that it doesn't fit well. So we use a more complex model (here a parabola) and get a much better fit.

We saw two ways to solve the problem: gradient descent vs. matrix algebra

Gradient descent updates the weights online, iteration over iteration over iteration. In matrix algebra, requires a matrix inverse operation which is very very expansive; for very large matrices (e.g. on Facebook with millions and millions of rows/columns) this will take very long.

NEW MATERIAL

Logistic regression: the "hammer" in machine learning, so-called because everyone uses it.

relatively fast;

parameters to tune are simple enough;

the problem is that "all of your problems are going to look like nails"

Example: student performance

gauge success or failure based on demographics, school district, parent's work, etc.

Equation

hb(x) = b0 + b1x1 + b2x2

In order to make this a classification (e.g. true/false), need to clamp hb(x) between two numbers. For example, 0 to 1.

Can do this mby setting a pivot value (called ReLU);

Or by setting a function f(x) capable of clamping the value -- this is preferred

Sigmoid function

g(x) = 1/(1+e^-z)

Defined for all real numbers, lim(z-->-inf)=0, lim(z-->+inf)=1

substitute the sigmoid function g(x) in for hb(x), with z = bx; now hb(x) is contained by (0,1).

This new function, in the car example, could tell us to buy the car if hb(x) is near 1, and to not buy it if hb(x) is near 0.

Optimization

Using the gradient descent method, this doesn't have a nice closed form.

So now think about a cost function that will map hb(x) back to a cost.

If we consider probabilities, we could say that the probability that y=1 (i.e. buy the car) is equal to the output of hb(x); conversely the prob that y=0 (don't buy the car) is 1-hb(x); this leads to the creation of our cost function, p(y|x,b)

Now use the maximum likelihood estimate method to determine the best values for b -- this is in the slides as L(b)

EXAMPLE: Chapter 3 regression/logistic regression

titanic passengers dataset, with 0/1 if the survived or not, their names, sex, age, etc.

Goal is to do logistic regression for good predictors of survival

First thing to do is clean the data -- taking out columns that contain any NaN values

Next we select only the fields that we are interested in: Pclass, sex, age, sibsp, parch, fare

In/Out[34] shows histogram of the passengers based on age

In/Out[35] shows distribution of sexes

In/Out[36] shos distribution of fares paid by passengers

In/Out[12] creates two new fields, 'male' and 'female', with 0 false and 1 true

In/Out[37]-[39] adds the male/female fields and removes the sex field

In/Out[42] assigns y as the field of those who survived from the original dataset, and takes the values in the table 'data' and puts it in the matrix x

In[43] does logistical regression on y using the predictors in x (see the top of the page for how the function 'clf' was defined.

Out[43] gives all the details of the logistic regression

[44] gives the weights for each of the fields in 'data'; negative value means inversely correlated, positive means positively correlated

[47] changes the penalty type -- we can go over penalties in the future