Notes

Xinyu Zhong
Wolfson College

February 5, 2023

Contents

Abstract

Abstract of this course

1 Introduction

mention supervised vs. unsupervised, supervised is more relavent for physics.

2 Overview of fitting

Although the idea of machine learning is first introduced in 1959, the most fundamental technique involved in machine learning dates way back in history when man explore ways to find the best method of curve fitting. Curve fitting is a process of construction a curve, or mathematics function, that has the best fit to a series of data points. It still remains as one of the most theoretically challenging part of machine learning.

Linear regression The most basic, and commonly seen fitting technique is a first order polynomial equation:

$$y = ax + b \tag{1}$$

which is a straight line that connects two points with distinct x coordinates. This is also known as linear regression.

Taylor Theorem With 3 data to fit, we could always add a term of higher power of x, to make it a quadratic equation

$$y = ax^2 + bx + c \tag{2}$$

or another term too construct a cubic regression:

$$y = ax^3 + bx^2 + cx + d \tag{3}$$

This reminds us a Taylor expansion only works for small x, disaster at large x. The limitation of this Taylor expansion comes when the x becomes an infinitely large value, which will cause the magnitude of y to become infinitely large, which many not reflects the datasets properly. Another limitation comes in when the number of independent variables becomes more than 1. For example, y is now a function of x_1 and x_2. i.e. $y(x_1, x_2)$. Taylor series cannot extrapolate the function for then the independet variable x becomes large In this case, we would have to include a term such as x_1x_2 and $x_1^2x_2$, which means that the number of coefficient we used is now grows exponentially to the number independent variable

Padé approximant A Pade approximant is an approximation of a function using rational polynomials. An $[N/M]$ Pade approximant is formed of a Nth degree polynomial on the numerator and an Mth degree polynomial on the denominator:

$$P(x) = \frac{a_0 + a_1x + a_2x^2 + ...a_Nx^N}{b_0 + b_1x + b_2x^2 + ...b_Mx^M} \tag{4}$$

This technique is developed by Henri Padé around year 1890. Padé approximant $\frac{ax^2+bx^3+...}{c+dx+...+x^6}$ making sure that $f(x)$ does not tend to infinity at large x, in this case tend to 1/x
Padé approximant does not have the same problem of using Padé approximant is superior to the Taylor series when describing function that contains poles. Also by dividing a polynomial by another, the Pade approximant prevents the function from diverging by letting $N <= M$

Neural network After 150 years or so
Neuro network, $\frac{x}{1+a|x|}$ using less indicator
after 30 years

Deep Neural network Deep neuro network, layers of sum of indicator the deepness refers to the layers

3 Different methods in Machine learning

Decision Trees Decision trees are used f two main types: classification tree and regression tree. We shall discuss the later in this article as in physics, we usually expect a numerical outcome Random forest/Gaussian etc.

Boosted Trees

Bootstrap aggregated

Random forest

4 Real life example

4.1 Tide-prediction Machine

William Thompson first invented