Notes

Xinyu Zhong
Wolfson College

February 22, 2023

Contents

Abstract

Abstract of this course

1 Introduction

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

2 Overview of fitting

2.1 Fitting techniques

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:

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

The objective is to minimise the ordinary least squares:

$$sumy_i - (kx_i + c)^2 \tag{4}$$

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:

$$P(x) = \frac{a_0 + a_1x + a_2x^2 + ...a_Nx^N}{b_0 + b_1x + b_2x^2 + ...b_Mx^M} \tag{5}$$

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 $N <= M$

2.2 Neural network

After 150 years or so
Neuro network, $\frac{x}{1+a|x|}$ using less indicator
Use a sum of such $\frac{x}{1+a|x|}$ element, which resembles a hyperbolic tangent. after 30 years
Neural Network is used to solve classification problems, sometimes regression problem. Example of Neural network code:

Back Propagation Backpropagation is just a way of propagating the total loss back into the neural network to know how much of the loss every node is responsible for, and subsequently updating the weights in a way that minimizes the loss by giving the nodes with higher error rates lower weights, and vice versa.

Gradient decent Slowly updating parameters

Activation functioon

1. Activation function: takes in a real vale and spit out a value between (0 to 1) to add linearity to the network.

2. Activation functions introduce an additional step at each layer during the forward propagation, but its computation is worth it. Here is why— In that case, every neuron will only be performing a linear transformation on the inputs using the weights and biases. It's because it doesn't matter how many hidden layers we attach in the neural network; all layers will behave in the same way because the composition of two linear functions is a linear function itself.

3. Different Types of activation function.

Loss function

Deep Neural network Deep neural network, layers of sum of indicator the deepness refers to the layers
Layers can be seen as $f(f'(f''(x)))$

3 Different methods in Machine learning

The most natural and commonly used machine learning technique is classification and regression

3.1 Gaussian Process

COvariant

3.2 Neural network

3.3 Deep learning

It requires very large amount of data in order to perform better than other techniques. It is extremely expensive to train due to complex data models.

3.4 Regression

3.5 Classification

The output is discrete, can be more than 2.
Classification does give step function
not differentiable.

Support Vector Regression Instead of minim sing the error, SVR give us the flexibility to define how much error is tolerable and will find an appropriate line. The term we are minimising is the efficient vector while we keep the error as a constraint.(Hypersurface) Non-linear Classification.

Decision Trees Decision trees are used for two main types: classification tree and regression tree. We shall discuss the latter in this article as in physics, we usually expect a numerical outcome.

Random forest/Gaussian etc.

Random forest

KNN K-nearest Neighbours

SVN Non-linear problem

4 Real life example

4.1 Gaussian Process

A rich article explaining a modified version of GP, which can mentioned previous researching in application of GP https://www.sciencedirect.com/science/article/pii/S0191261521000369

4.2 Gravitational Waves

Problem metric that doesn't changed when shift or coefficient shift-free and dimension free R^2

r^2 different things.

5 Choosing ML techniques

Testing protocal Train validate test 5 fold. Tranlational text recogonisation.

Take fouriour transform for symmetery Translationall invariant.

not invariant forest, invariant covolution neural network.

Gaussian more perfect,

6 Conclusion

Put exam bullet point in conclusion. Also, show future prospect, big advancement, is physics ready machine learning, what is the data collection like in physics

MCC for categorical data.

Example: Overall structure of this review.The over structure should be general concepts about machine learning followed by specific example.

Example: How deep should I look into each case. For example, different activation function for neural network, different definition of distance in KNN problem, different definition of Loss function in liner regression?

Example: Marking Scheme:communication/ Would they read my report, if yes, methods/how to next step

Sum of F(ax+b)+F(ax+B) more superior to the Parday when no fit data with lesser parameters. independently tested.