Overview of different Gradient Descent Optimization technics

David Panou

October 1, 2016

1 Introduction

Statistical Modelling has been seperated into two distincts communities : the *data modelling* community, filled with statisticians and the *algorithmic modelling* community that is mostly made out of young computer scientists that had to model a stochastic. Throughout this paper, I will go in depth about the differences of those two cultures and show how data modelling has kept most of the statisticians community to work on exciting new problems.

2 Statistical Modelling

A statistical modelling problem is one on which we are facing a natural phenomenon that we wish to either *explain* or *predict*. Given that definition the problem can modelled using the following schema :

Figure 1: A natural phenomenon can be represented by a flow chart with a black box

There are two common ways to approach a statistical modelling problem :

2.1 Data Modelling

On data modelling, we first create a stochastic model that imitates the behaviour of the black box model represented on **??** can be filled using the schema drawn on **??**

Figure 2: A data modelling flow chart

Data modelling has been. Newton's point physics is a good data model, that respects Occam's Razor. It is good model of how the mechanic works on earth and is still widely used now.

However, data modelling has some concerns that prevents todays statisticians to get involve in a whole new variety of problems that have to be solved with todays technology.

1

2.2 Algorithmic modelling

Algorithmic Modelling has emerged through communities other than statisticians, mostly young computer scientist. This approach is also sometimes called machine learning.

An Algorithmic Modelling approach focuses on optimizing models, considering that the nature mechanism producing the output is totally unknown. It can be represented by the following scheme :

Figure 3: A algorithmic modelling flow chart

This type of modelling is usually evaluated through accuracy and trained error. The better the accuracy, the better the model is.

3 Problems that Data Modelling faces and how algorithmic modelling solves them

One major reminder that we have to keep in mind while dealing with data models is that *conclusions drew, are conclusions about the model mechanisms and not the natural phenomenon mechanisms.* So the results of the data models is dependent on how good the model represents nature.

Every article about data modelling started with the same sentence : "Assume that the data are generated by the following model : ..."

The issue with those assumption are that a single model can't always encompass a natural phenomenon. In some cases, it is questionnable that a linear model can even fit the data. However, data models have been used extensively by the statistics communities

Nevertheless problems come with data models, and I named them the following :

- **Rashomon** : the multiplicity of good models

- **Occam** : simplicity over complexity

- **Curse of dimensionality** : cursing or blessing?

3.1 Rashomon : The multiplicity of good models

When using data models, there is a assumption is that the model fit the data. In order to gain information and to have a more robust models (i.e. models that have less variance) we can reduce the dimensionality of our input and select a subset of parameters that explain our the most. This can be achieve by subset selection in linear regression for example.

But most of the time, there is more than one subset with k variables that fit the data and the question is now which one do we select?

3.2 Occam : simplicity over complexity

Occam's razor stated merely that "entities" (or explanations) should not be multiplied beyond necessity [**?**] , and this usually implies choosing simpler models that are interpretable though have lower accuracy rather than complicated model with good accuracy but are hard to interpret.

Data modelling usually go in favour of the Occam Razor, to favour interpretability, but better insight about the underlying mechanism can be obtain by good predictive accuracy that actually extract deeper connections between input variables.

We have to remember that the goal of statistical modelling is accurate information and not interpretability. This can be achieved with good accuracy.

3.3 The curse of dimensionality

When analysing a given problem, there is sometime a issue about the dimensionality of the input variables being too large. A common way to deal with this issue in data modelling , is to reduce the dimensionality of of the input variable by selecting the most meaningful ones. But reducing dimensionality reduces the amount of available information.

Going the opposite direction, algorithmic modelling usually combines input variables thus projecting them in a larger (sometimes infinite) dimension.

4 Conclusion

Data modelling has been used a lot and has proven to be effective to certains points (see Newton's mechanic). But in today's world, most simple problem have already been addressed and more pragmatic method need to be used to gain information about involving terabytes of data.

References

[1] Richard O. Duda, Peter Hart, David G. Stork *Pattern Classification.*

[2] Leo Breiman. *Statistical Modeling: The Two Cultures.* Statistical Science : Vol. 16, No. 3, 199–23, 2001