PolyChord

Technical notes for Commercial Users

Mike Handley

January 25, 2018

1 What it is, who it is for and what it does

PolyChord is a piece of data science software. It fits models to data, acting as an alternative to optimisation tools or Markov-Chain Monte-Carlo approaches. PolyChord is a software tool for use by data scientists in commercial, industrial and scientific research and development departments. A generalised data science problem can be broken down into stages:

1. Gather and curate your data.

2. Construct models (or refine existing ones) for describing your data.

3. Fit/train these models.

4. Select the best model.

5. Use the model to make predictions.

PolyChord provides a cutting-edge solution to steps 3 and 4. PolyChord represents the cutting-edge in nested sampling. It fits models to data using a Bayesian-inspired sampling approach, and allows for model comparison by computing the marginalised likelihood (Bayesian Evidence).

2 The basis on which PolyChord works

Numerical modelling (in a broad sense) is the integral part of how we understand and use data. In order to extract information from data, one first constructs a model. One then 'fits' or 'trains' the model before using the model to extract information about the data, and make further predictions. Modelling acts to compress big data into a manageable and usable tool. Models come in many flavours, the two broad classes being Generative and Discriminative. Scientific models tend to be generative, whilst machine learning models (such as neural networks) tend to be discriminative. Both types however require model fitting or training. PolyChord has competitive advantages in both of these areas. PolyChord is a black-box optimisation tool, as it self-tunes and requires minimal

user intervention. PolyChord represents the cutting edge of evidence computation for model comparison, providing more accurate and reliable answers than all existing tools in an achievable computational timescale. PolyChord does not require gradients, merely the optimisation function to be explored, and the region to explore it in. PolyChord is supported by several of the world-experts in the field of nested sampling.

3 How PolyChord compares with other available tools.

Traditionally model fitting is performed by a variety of approaches. These fall into two categories: Maximisers and Samplers.

3.1 Maximisers

Maximisers fit models by finding the single 'best fit' parameters of the model. Examples:

- Gradient methods

- Stochastic optimisation

- Genetic algorithms

- Derivative-free approaches (e.g. BOBYQA)

These approaches are typically the most performant, as they rapidly ascend to a peak of function to be optimised. They can however get trapped in local minima, and it is in general very challenging using these methods to evaluate whether or not this peak is the best, or even unique. Maximisers also suffer from problems with regularisation and the combatting of over-fitting.

3.2 Samplers

Samplers fit models by finding a collection of 'most typical' parameters of the model. Unless the model is particularly simple, this is usually performed by Markov-Chain Monte-Carlo (MCMC) approaches:

- Metropolis Hastings

- Gibbs Sampling

- Hamiltonian Monte Carlo

- Ensemble sampling

- Simulated Annealing

- Sequential Monte Carlo

- Nested Sampling

Sampling a function naturally combats over-fitting, allowing automatic quantification of the errors in your analysis and the fidelity of your fit. We prefer sampling to maximising. A fast unreliable answer is in general no-where near as useful as a more careful correct one. Typically what we do when exploring an unknown distribution is to start by using PolyChord, and then once the problem is well understood, (if necessary) switch to a more carefully tuned optimiser. Nested sampling is unique in this field in its ability to navigate complex optimisation functions in high dimensions, with features such as multiple modes and curving degeneracies. PolyChord is the cutting edge of nested sampling.

4 How does PolyChord compare to MultiNest?

If you've come across nested sampling before, then you have likely heard of MultiNest. PolyChord and MultiNest were both created in the Cavendish Laboratory and share co-creators. As the first performant implementation of nested sampling, MultiNest has been widely adopted by the research and industrial community, across a wide variety of fields. However, MultiNest was only ever meant to be a first step in the chain of nested sampling algorithms. As it functions using an advanced rejection-sampling approach it performs optimally in low dimensions, but has an exponential scaling as the dimensionality of the problem is increased. PolyChord represents the next step in nested sampling, with polynomial dimensionality scaling (quadratic for maximally hard problems, sub-linear for realistic ones). It is competitive with MultiNest in low dimensions, and far superior in higher dimensions How would my organisation work with PolyChord? You need our team to sit down with your team and decide what models you want to fit and train - and so determine what outcomes you want to get from PolyChord. We do this at a management and technical level so everyone is happy with the agreed targets. We then run a technical audit aimed at uncovering how you are currently gathering and curating data, how your current models are built and a quick examination of fitting and training protocols. We then discuss and specify a bridging tool and planned outputs. In nearly all cases, a tool customized to meet your particular requirement is needed as everyone is using and storing their big and complex data in different ways

5 Using and deploying the tool

In nearly all cases, companies and organisations want PolyChord to run on their own premises. We have means of doing this. In larger more long-running projects we can work on projects from secure facilities within our Cambridge base.