Hi Nick,

The next task in our ML project is parameter tuning. Please use Bayesian Optimization – a strategy for both supervised and unsupervised parameter tuning. It aims to both explore and exploit the parameter space.

To understand BO, imagine you are in a boat on the ocean looking for calm seas. Initially, you'll have some belief about where to go, so you'll head in that direction. When you arrive, take note of your location and the sea around you. Then repeat this process several times, recording all information on each iteration. Eventually, you'll have a good idea of where calm waters are. This is what B.O. emulates.

Naïve implementations, like grid search and random search, could be used. Think of grid search as sailing the ocean methodically until you've been *everywhere* – a laborious task. You could also point the ship in a random direction and hope for the best: random search. Each method may lead to the same place eventually, but only Bayesian Optimization uses the knowledge available to make an educated decision about where to go next.

The key components of Bayesian Optimization are:

1) **Surrogate Model**: This approximates the true objective function. If the true objective function is "The Ocean," then your experience on the ocean trained your mental map of the Ocean, which in this case is the surrogate model. We use the surrogate model in place of the objective function if we don't have the true objective function or if it's prohibitively expensive to compute. In practice, Gaussian Processes are a common choice for the surrogate model because they provide predictions and uncertainty and are inexpensive to update.
2) **Acquisition Function**: When deciding where to go next, you'll have several locations in mind. The acquisition function decides which of these to head in. It needs to balance exploration (highly unpredictable areas) with exploitation (areas known to be desirable).

Please construct an approximation to the true objective function. One approach is to manually label a set of data and evaluate the output of our model. Another approach is to use the silhouette score – a metric that evaluates the clustering configuration. Each record receives a score based on how similar it is to its cluster. If most records have a high score, the model is performing appropriately.

Then make initial guesses for the parameters k and t. Run the model with these guesses and calculate each using the approximate objective function. Then use these (**x**, y) samples to train the Gaussian Process, which will model the relationship between the hyperparameters (**x**) and the objective function score (y), providing predictions and estimates of uncertainty for any set of parameters.

An acquisition function can be used to decide where to sample from next. Existing libraries can be used for its implementation. Sample and repeat until the stopping criterion is met. Each (**x**, y) sample procured should be used to update the surrogate model. Finally, choose the parameters that gave the best performance according to the objective function.

Regards,

Michael