Cross-validation based adaptive sampling for Gaussian process models

When we are working with complex computer codes we want to able to perform sensitivity and uncertainty analyses as cheaply as possible. Gaussian process (GP) emulators are often used to approximate the output of the expensive code. We consider the problem of extending a space filling initial design to sequentially improve the emulator. Sequential sampling has many advantages over a ‘single shot’ design as we can learn about the space from previous model runs and place future design points where they can have most effect. A sequential sampling approach based on leave-one-out (LOO) cross-validation is proposed that can be easily extended to a batch mode. This is a desirable property since it saves the user time when parallel computing is available. After fitting a GP to training data points, the expected squared LOO error (ESE-LOO) is calculated at each design point. ESE-LOO is used as a measure to identify important data points. More precisely, when this quantity is large at a point it means that the quality of prediction depends a great deal on that point and adding more samples in the nearby region could improve the accuracy of the GP model. As a result, it is reasonable to select the next sample where ESE-LOO is maximum. However, such quantity is only known at the design points and needs to be estimated at unobserved points. To do this, a second GP is fitted to the ESE-LOO s and where the maximum of the modified expected improvement (EI) criterion occurs is chosen as the next sample. EI is a popular acquisition function in Bayesian optimisation and is used to trade-off between local/global search. However, it has tendency towards exploitation, meaning that its maximum is close to the (current) “best” sample. To avoid clustering, a modified version of EI, called pseudo expected improvement (PEI), is employed which is more explorative than EI and allows us to discover unexplored regions. The results show that the proposed sampling method is promising, outperforming existing methods.

This is joint work with Hossein Mohammadi