Active Learning for Level Set Estimation

Alkis Gotovos, Nathalie Casati, Gregory Hitz and Andreas Krause
ETH Zurich

Problem

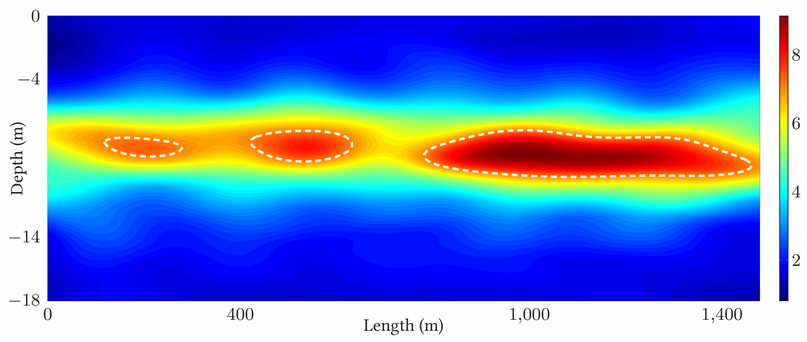
We would like to determine the regions where the value of some unknown function lies above or below a given threshold level.

The above can be posed as a classification problem (into super- and sublevel sets) with *sequential* measurements, which are assumed to be *expensive* and *noisy*.

Example applications

Environmental monitoring

Estimate regions of a lake transect where algae concentration is “abnormally high”.



Geolocating internet latency

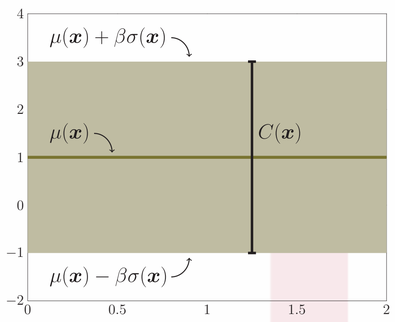
Estimate regions of the world with “acceptable” latency to our PC, e.g. for trouble-free online gaming.

Gaussian processes

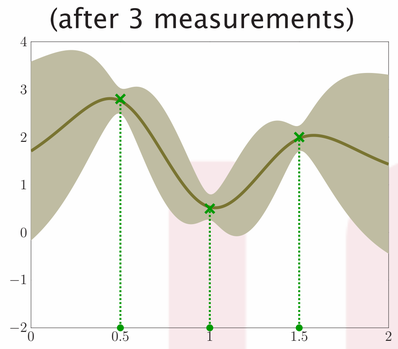
Estimation

Given some measurements, GPs provide *mean and variance* estimates of the unknown function, allowing us to construct *confidence intervals* at each point.

Prior

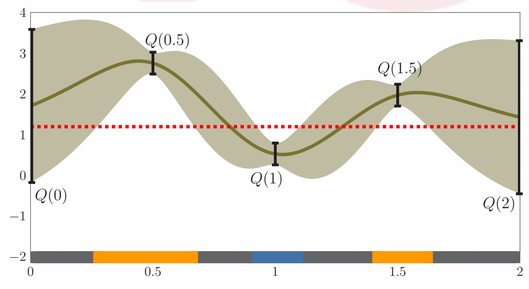


Posterior
(after 3 measurements)



Classification

For each point, we use the GP-derived confidence intervals to either classify it into the super- or sublevel sets, or leave it unclassified.



Measurement selection

To obtain informative measurements w.r.t. the problem at hand, at each iteration we select the most *ambiguous* point among the yet unclassified to be measured.

Intuitively, ambiguity quantifies our difficulty in classifying a point w.r.t. the given threshold level.