Introduction to Gaussian Processes

Part 2: Classification

Ruben Martinez-Cantin

Defense University Center
Zaragoza, Spain
rmcantin@unizar.es

Outline

What you will see here:

- Gaussian process hyperparameters
- Regression
- Binary classification
- Active learning and experimental design
- Submodularity
- Bayesian optimization
- Stochastic bandits

Classification as regression

- We have seen that GPs are suitable for regression
  - Can we use then for classification?
- Regression for classification: an classic idea in machine learning.
  - Idea: Instead of predicting the labels y, we predict the likelihood of the label $p(y|x)$.

$$p(y = +1 \mid \mathbf{x}) = \lambda(\phi(\mathbf{x})^T \mathbf{w})$$

.

- Our likelihood function needs to be a map $f : \mathcal{X} \rightarrow [0, 1]$
  - Logistic regression

$$\lambda(x) = \frac{1}{1 - e^x}$$

  - Probit regression.

$$\lambda(x) = \Phi(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

Classification with GPs

- GPs gives as a distribution over functions $p(f_*|x_*, \mathbf{x}, y)$.
- It should be straightforward to infer a label

$$p(y_* = +1|x_*, \mathbf{x}, y) = \int \lambda(f_*) p(f_*|x_*, \mathbf{x}, y) \, df_*$$

- But this time we need to solve this integral before

$$p(f_*|x_*, \mathbf{x}, y) = \int p(f_*|x_*, \mathbf{x}, f) p(f|\mathbf{x}, y) \, df$$

  where $p(f|\mathbf{x}, y) = p(y|f)p(f|\mathbf{x})/p(y|\mathbf{x})$ is the posterior over the *latent parameters*.
- Where is the problem?

Problem

- When we have a Gaussian likelihood everything is nice.
  - Exact solution.
  - Closed-form.
  - Highly efficient computation.
  - Easy extension of the model, hierarchical representations ...
- For classification we cannot use a Gaussian likelihood.
- Logit and probit likelihoods do not admit closed form.
  - We need to compute an approximation of the posterior $p(f|\mathbf{x}, y)$
  - Laplace approximation, Expectation Propagation, MCMC, ...

Handwritten digit classification