CS4070: EXERCISE 3: GAUSSIAN PROCESS CLASSIFICATION

Hand in before January 16 (2020), 16.00 (in the postbox at the first floor of the Van Mourik Broekmanweg.)

(1) (a) Write a script that samples from a Gaussian process $f : \mathbb{R} \to \mathbb{R}$, for given mean-functions μ and kernel K.

(b) Consider the following 3 kernel functions: square-exponential kernel, polynomial kernel of degree 2, neuronal kernel.
For each of these kernel functions, draw 5 realisations from the Gaussian process and overlay these in one plot. Consider $x \in [-5, 5]$. Play around a bit with the hyperparameters in these kernels so that "typical" paths are displayed. Include the plots in your report. Note that for numerical stability you may wish to add a small multiple of the identity matrix to the Gram-matrix (which is the matrix with elements $K_{ij} = K(x_i, x_j)$).

(2) *This exercise refers to the final 6 slides in the handout on Gaussian process regression (Gibbs sampling for GP classification).*

Assume data $\{(x_i, y_i),\ i = 1, \ldots, n\}$ where $y_i \in \{0, 1\}$. Let X denote the design matrix for this problem. Assume the model

$$\begin{aligned} y_1, \ldots, y_n \mid f, X &\overset{\text{ind}}{\sim} Ber\left(\Phi\left(f(x_i)\right)\right) \\ f &\sim GP(0, K) \end{aligned}$$

where Φ is the cumulative distribution function of the standard normal distribution.

Assume

$$K(x, \tilde{x}) = \exp\left(-\frac{1}{2\ell^2}\|x - \tilde{x}\|^2\right) + 10^{-5}\mathbf{1}_{\{x=\tilde{x}\}}$$

for length-scale ℓ.

(a) By conjugacy, the update step for $(z_1, \ldots, z_n)$ amounts to drawing from the multivariate normal distribution. Derive its parameters.

(b) Implement the Gibbs sampler that iteratively updates $\{f_i\}_{i=1}^n$ and $\{z_i\}_{i=1}^n$.

(c) Augment the Gibbs sampler by adding the parameter v in the model, such that instead of $f \sim GP(0, K)$ we assume $f \sim GP(0, vK)$. Assume apriori $v \sim \text{InvGamma}(2, 2)$. Hence, in this Gibbs sampler, the steps implemented under *(b)* are conditional on v, and there is one extra step in which v is updated.

Date: December 27, 2019.

(d) Make a small test dataset to verify that the chain "converges", i.e. the traceplots settle to a stationary regime. Include such traceplots in your report (you could for example make traceplots of a few z_i, f_i coefficients as well as v). For constructing a test set, you can *for example* take x one-dimensional and sample x-values uniformly on $[-1, 1]$, subsequently choose some f and sample corresponding y-values.

(e) Consider the data in `train.csv`. Here, the column named `y` is the response y; the remaining columns are predictors. Set the length-scale according to the rule of thumb

$$\ell^2 = \frac{dn}{\sum_{i=1}^n \|x_i\|^2},$$

which is the number of predictors divided by the mean square Euclidean norm of the predictors. Let $x^\star$ denote a row from the data in `testX.csv`. Generate samples from $p^\star = \mathbb{P}(y^\star = 1 \mid x^\star, X, y)$. Please refer to steps A-B-C in the handout on prediction under Gaussian process classification. For rows 1, 2 and 17 in `testX.csv` display a histogram of these samples.