Statistical Methods 2

Portfolio 7 – Gaussian process regression

Complete the following tasks and submit your work on Blackboard by 4pm on Friday 21/04/2023

Task 1 (30 marks)

Show the following results:

1. For any function $g : \mathcal{X} \to \mathbb{R}$, the function $k(x, x') = g(x)g(x')$ is a kernel.
2. For any constant $a \geq 0$, the function $k(x, x') = a$ is a kernel.
3. For any $m \in \mathbb{N}$, kernels $\{k_j\}_{j=1}^m$ and non-negative real numbers $\{c_j\}_{j=1}^m$, the function $k = \sum_{j=1}^m c_j k_j$ is a kernel.
4. If k is a kernel on $\mathbb{R}^p$ and $\mathcal{X} \subset \mathbb{R}^p$ then the restriction of k to $\mathcal{X} \times \mathcal{X}$ is a kernel, i.e. the function $k(x, x')\mathbf{1}_{\mathcal{X}\times\mathcal{X}}(x, x')$ is a kernel on $\mathcal{X}$.
5. (**Optional**) For any $m \in \mathbb{N}$, kernels $\{k_j\}_{j=1}^m$ and non-negative real numbers $\{c_j\}_{j=1}^m$, the function $k = \prod_{j=1}^m c_j k_j$ is a kernel.

   Hint: Recall that for a real symmetric positive semi-definite matrix $\boldsymbol{M}$, the eigendecomposition of $\boldsymbol{M}$ gives a matrix $\boldsymbol{A}$ such that $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{A}^\top$.

Remark: In Gaussian process regression, the above properties of kernel functions are sometimes used to build a “good” kernel k for the GP prior (see Rasmusen’s book, Section 4.2.4 and the example of Section 5.4.3).

Task 2 (70 marks)

Choose a dataset $\{(y_i^0, x_i^0)\}_{i=1}^n$, where $x_i^0 \in \mathbb{R}$, on which you can fit a Gaussian process regression model with known variance $\sigma^2 = \lambda$.

Then,

- Choose a kernel k with parameter ψ.
- Compute exactly the posterior distribution of f given the observations $y_{1:n}^0$ (if n is large do that for a subsample of your dataset). More precisely,
  1. Choose λ and at least one component of ψ using the empirical Bayes approach.
  2. Make a plot showing the function f_n, the corresponding credible sets at level 95% and the observations $\{(y_i^0, x_i^0)\}_{i=1}^n$
- Repeat 1. and 2. when the posterior distribution is approximated using the low rank approximation approach discussed in Chapter 12, with $m < n$. Assess the sensitivity of your results to the choice of the set $\{\tilde{x}_i^0\}_{i=1}^m$.

For this task you can for instance use the same dataset as in Portfolio 5 (Task 2), or be more ambitious and try a dataset with $p > 1$ predictor (e.g. the `ozone` dataset used in Chapter 9) . To the best of my knowledge there exists no `R` package which implement Gaussian process regression using the empirical Bayes approach to choose the hyper-parameters. To implement the empirical Bayes approach you can write a function that compute the marginal log-likelihood and then optimize it using `optim`. Be careful, maximizing the marginal log-likelihood is not a convex optimization problem. It is therefore a goof idea to perform the optimization for different starting values.