Master M2 MVA 2021/2022 - Introduction to (deep) Probabilistic Graphical Models - Homework 2

These exercises are due on or before January 14th 2022 and should be submitted on the drop box available online. They can be done in groups of two students. The write-up can be in English or in French. Please submit your answers as a (unique) zip file that you will name MVA DM2 <your name>.zip if you worked alone or MVA DM2 <name1> <name2>.zip with both of your names if you worked as a group of two. Indicate your name(s) as well in the documents. Note that the zip file should weight no more than 16Mb. Only such zip files with such names will be evaluated. Your solutions as well as your codes should be present in the zip file. Again, we recommend you to write your code and report thanks to a notebook or Markdown file (in R or Python for the first exercise, and in Python for the second one).

1 - Gaussian processes

We consider a Gaussian process model (GP) for regression. Thus, denoting Y the set of all (n) outputs (in $\mathbb{R}$) and X the feature matrix of size $n \times p$, the model gives:

$$Y|X \sim \mathcal{N}(0_n, C),$$

where the elements of the covariance matrix C are such that:

$$C_{ij} = k(x_i, x_j) + \sigma^2 \delta_{ij}.$$

The function δ_{ij} is 1 if $i = j$, 0 otherwise. Moreover, $k(x_i, x_j)$ denotes the kernel function between observations i and j. In this exercise, we consider the exponential quadratic kernel:

$$k(x_i, x_j) = \theta_0 \exp\{-\frac{\theta_1}{2}||x_i - x_j||^2\} + \theta_2 + \theta_3 x_i^\intercal x_j.$$

Finally, $\theta = (\theta_0, \theta_1, \theta_2, \theta_3, \sigma^2)$ denotes the set of all (hyper)parameters with σ^2 characterizing the variance of the noise.

1. Learning

Write a complete set of functions in R or Python implementing the training GP step, i.e. the optimization in θ of the (marginal) log-likelihood:

$$h(\theta) = \log p(Y|X, \theta).$$

2. Prediction

Write a complete set of functions in R or Python implementing the prediction GP step, i.e. the prediction $\hat{y}$ for a new observation x.

3. Real data

Learn the GP regression model on the first 75% part of the **UScrimes.csv** data set. The target is the Murder variable. Use the model learned to predict the outputs of the remaining 25%. Compare the RSME you obtain with the RMSE of a simple multivariate linear model for regression.

For the linear model, you can use existing packages. However, all the functions you implement should be stand alone. You are allowed to rely on packages but not on the ones implementing directly solutions to the questions. The source code should be handed in along with results.

2 - IWAE

The goal is to train a deep latent variable model on a binarised version of MNIST, using tensorflow. Details are provised in the companion notebook.

1. Ancestral sampling

Implement a function that performs ancestral sampling for this deep latent variable model, and show 5 sampled images from the initialised model. Of course, we expect these samples to look like random noise, since we have not trained our model yet.

2. Training

Create a function that computes an unbiased estimate of the IWAE bound, and optimise it to train the model. You are allowed to use TF probability to avoid implementing the reparametrisation trick, and you are allowed to use keras for training. What are the values of the train/test IWAE bounds after training?

3. Sampling

Show a few samples from the model after training.

4. Importance sampling and visualisation

Implement a importance sampling estimate to approximate the posterior mean of each data point:

$$\mathbb{E}[\mathbf{z}|\mathbf{x}_i] = \int \mathbf{z}p(\mathbf{z}|\mathbf{x}_i)d\mathbf{z}.$$

Visualise these two-dimensional embeddings using a scatter plot.