Reports for the three assignements (Robust deep learning) - MVA 2022/2023

Dorian Gailhard dorian.gailhard@telecom-paris.fr

March 28, 2023

1 Assignment 1 - Bayesian Linear Regression

1.1 Question 1.4

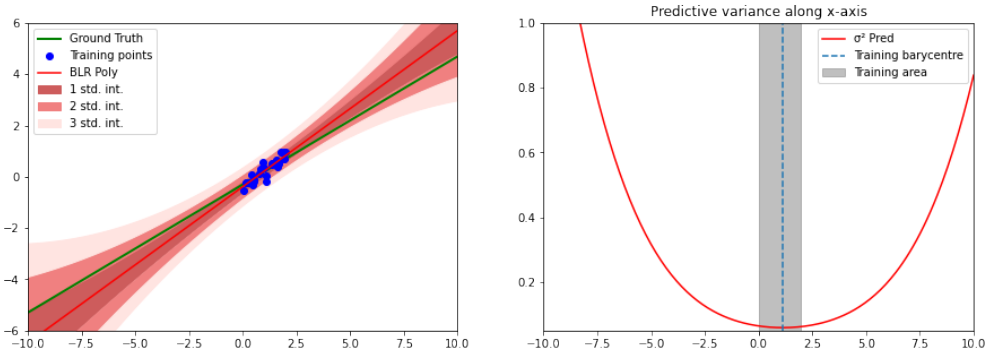

Figure 1: Results obtained with the first design matrix

1.2 Question 1.5

Intuitively, as this is a linear regression, a small variation of the slope causes a slight difference at the origin, but cause an increasing divergence as we get far from it. At the origin, only the intercept (parameter b in $y = ax + b$) has an effect and while the slop becomes increasingly prominent the farther we get.

Here with $\beta = 1$ and $\alpha = 0$, the value of the variance is $1 + \Phi(x)^T \Sigma \Phi(x)$ with $\Sigma = (\Phi(\mathscr{D})^T \Phi(\mathscr{D}))^{-1}$.

As $\Phi(x) = (1, x)$, when $\|x\| \to +\infty$ then $1 + \Phi(x)^T \Sigma \Phi(x) \to +\infty$.

1.3 Question 2.4

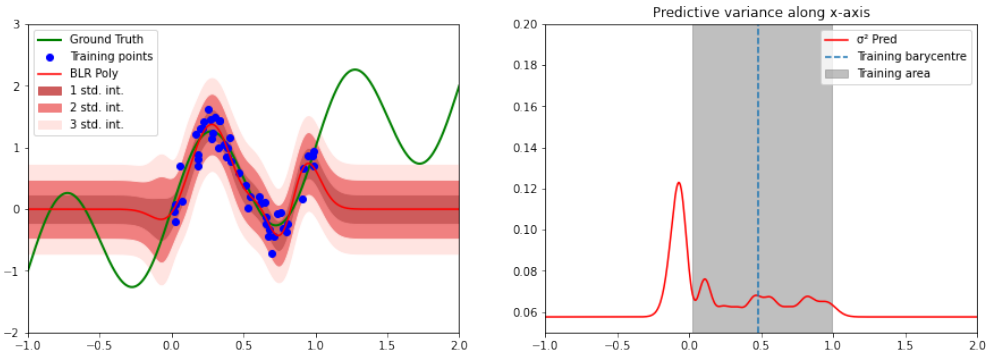

Figure 2: Results obtained with the gaussian design matrix

1.4 Question 2.5

When we get far from the training area, all the ϕ_j converge to 0 and the only remaining term in the variance value is β^{-1}, which is a constant.

2 Assignment 2 - Approximate Inference in Classification

2.1 Question 1.2

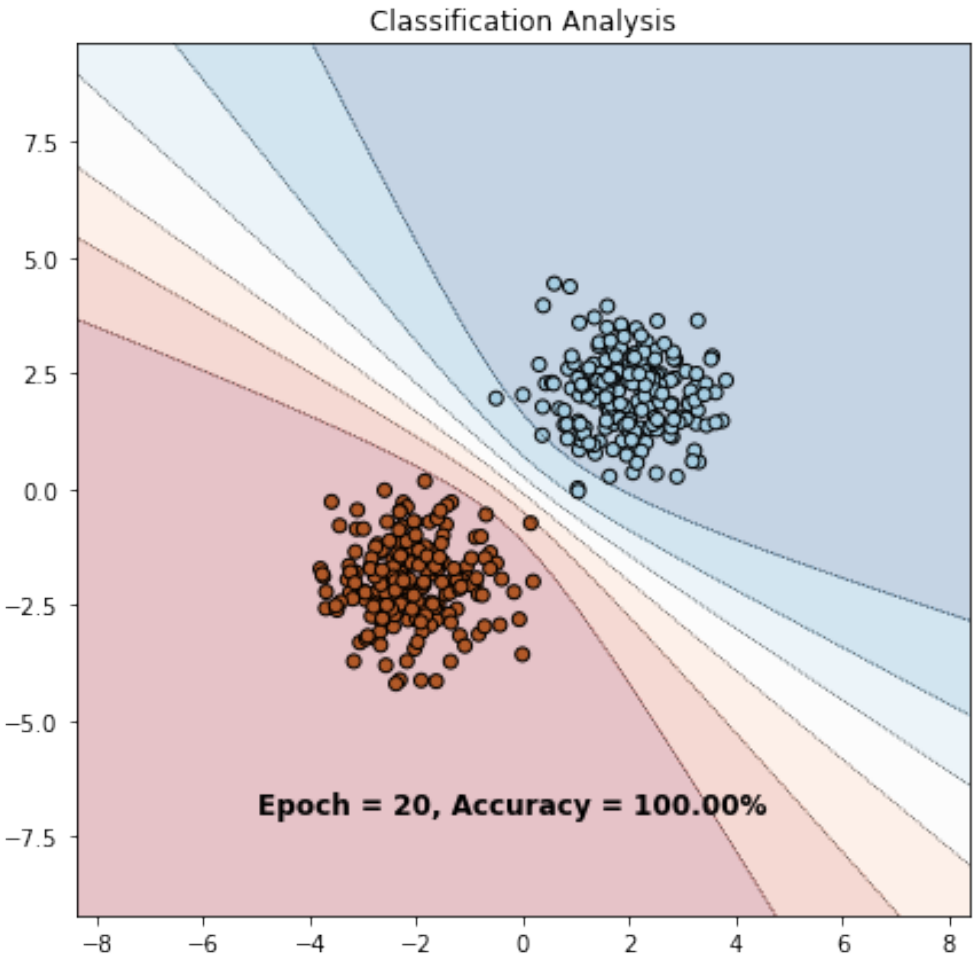

Figure 3: Results obtained with the Laplace approximation

This time it looks much more like the real distribution, the points far from training data have an increasing uncertainty with the distance, even if the landscape of uncertainty is a bit different from the true one.

2.2 Question 1.3

Laplace approximation approximates the real distribution by a unique gaussian distribution, whereas variational inference approximate the real distribution by complex operations on multiple gaussians distribution which are learned. This looks similar to a gaussian mixture approximation except that we have a transfer function.

2.3 MC dropout

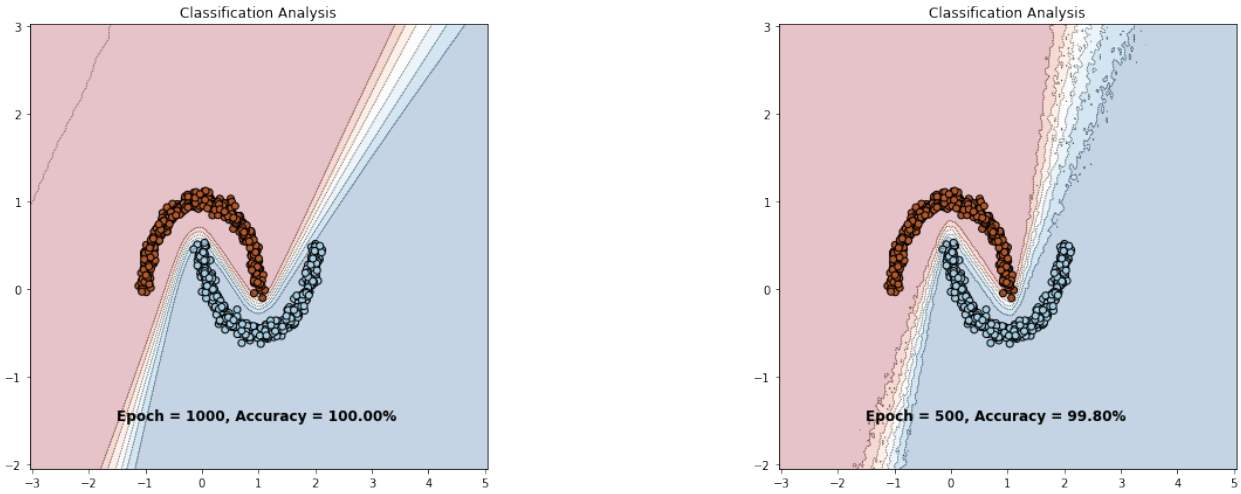

Figure 4: Comparison of variational inference and dropout

The distribution given by dropout is much more noisy than the distribution given by variational inference.

3 Assignment 3 - Uncertainty Applications

3.1 Question 1

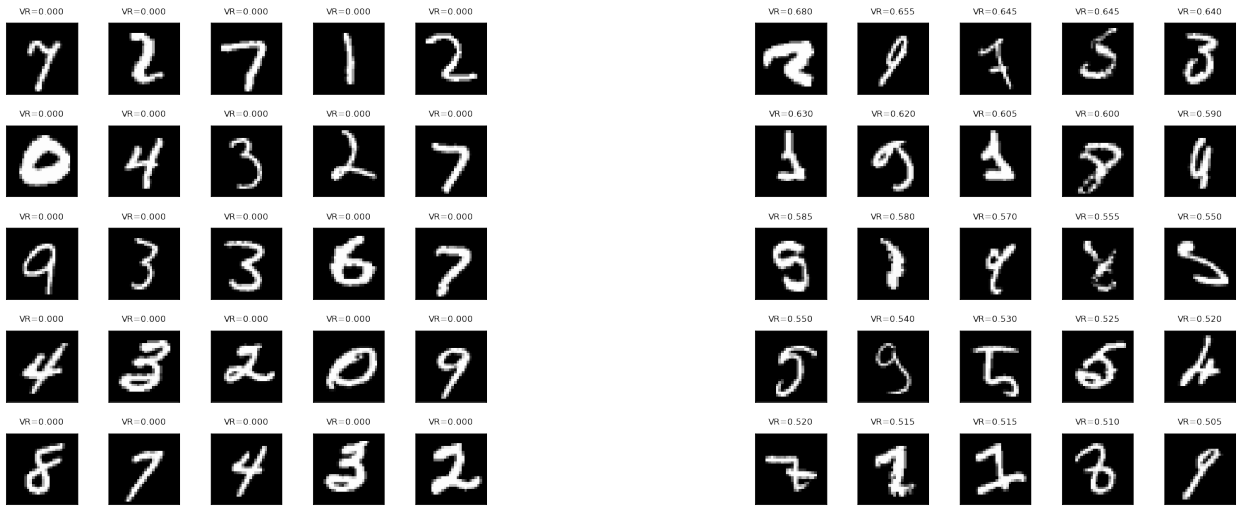

Figure 5: Digits associated to high and low confidence

Indeed, the confidence metric follows what we intuitively think : a var ratio of 0 means there is no ambiguity while a varratio of 0.5 means the digits are not so clean and it is not so clear what they are.

3.2 Question 2

Failure prediction tries to quantify the confidence of a model in its predictions, so that when the confidence is too low it can just output an error and select a fall-back action instead of outputting a false prediction.

LeNetConfidNet model is a simple LeNet with frozen convolutional layers that are shared with the classification model. It learns to output the true class probability of the sample.

The results are strange, my implementation is faulty.

3.3 Question 3

The results are strange, my implementation is faulty.

MCP uses the class probabilities output by the model, while dropout computes the confidence by hiding different parts of the network, ie it looks if all the parts of the model are in accordance or if it the model is split between different classes. ODIN tries to make the model robust to adversarial changes, ie to make the model probabilities output more robust to small variations, so that a difference in probabilities means a difference in the images.