Dropout as a Bayesian Approximation

ICML 2016

Summary

In this paper, the authors propose a new technique called *Monte Carlo Dropout* that re-uses information contained in conventional dropout neural networks to approximate predictive uncertainty. The algorithm is based on a link between gaussian processes and dropout networks and allows for the extraction of uncertainty information without an increase in computational complexity or a decrease in classification accuracy. Monte Carlo Dropout is shown to provide reliable uncertainty estimates on regression, classification and reinforcement learning problems.

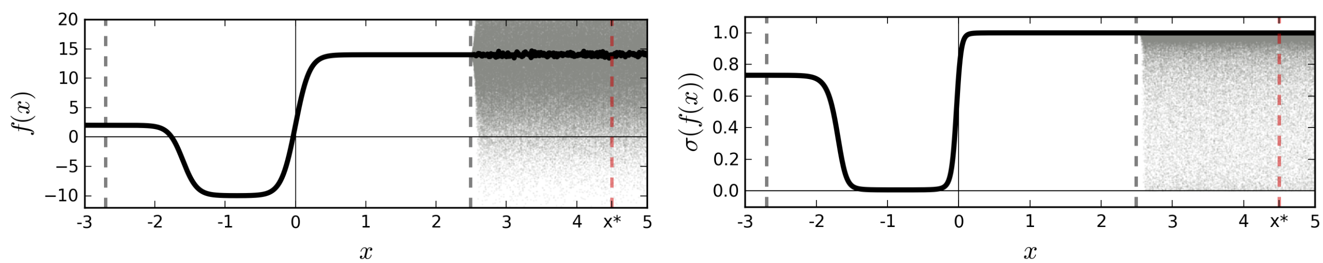

Figure 1: Illustration of input and output of a softmax function. Training data was used between the dashed gray lines. We can observe that the output of the softmax function on out-of-distribution data is much smoother than the input. A point x^* is consequently classified as class 1 with probability 1.0 despite a very noisy input function around that point. Uncertainty estimates could indicate the unfamiliarity of the network with points around x^*.

Main Contributions

- **Dropout and Gaussian Processes** The authors put forward a mathematical derivation of the link between dropout and gaussian processes that allows for a probabilistic interpretation of dropout. The paper shows that the objective of dropout corresponds to minimizing the KL divergence between a variational distribution and the posterior of a gaussian process.

- **Monte Carlo Dropout** A method is put forward that allows for the extraction of uncertainty information from standard dropout networks by applying dropout not only during training but also during inference. Thus, uncertainty information can be obtained without any additional computational cost commonly associated with probabilistic neural networks.

- **Classification Performance** The performance of MC Dropout is evaluated on MNIST, showing how the obtained uncertainty measures create robustness in the classification of rotated digits and allow for more interpretable results.

Implementation Details

- **Variational Inference** In order to approximate the intractable posterior over the network's weights, a variational distribution is selected that resembles the dropout approach by randomly disabling weights according to binary values sampled from a Bernoulli distribution:

$$\mathbf{W}_i = \mathbf{M}_i \cdot \mathrm{diag}([\mathbf{z}_{i,j}]_{j=1}^{K_i})$$
$$\mathbf{z}_{i,j} \sim \mathrm{Bernoulli}(p_i) \text{ for } i = 1, ..., L, \ j = 1, ..., K_{i-1}$$

- **Model Uncertainty** As mentioned before, the model uncertainty can be obtained from networks trained with dropout using standard point-estimates of the network parameters. The output is then defined by the mean and the variance computed over several forward passes using the same dropout scheme that was used during training. The authors mention that the predictive distribution $q(\mathbf{y}^*|\mathbf{x}^*)$ is expected to be of multi-modal nature due to the underlying assumptions of the variational approximation. Consequently, the characterization by mean and variance can only give an idea of the full distribution's characteristics.

Evaluation

- **MNIST Classification** The evaluation is conducted on a LeNet architecture trained on the full MNIST dataset. During training dropout is applied before the last fully-connected layer and the dropout percentage is set to 0.5. The network is presented with a continuously rotated version of the digit 1 and 100 forward passes are evaluated per instance. The results of the experiment are displayed in Figure 2. We can observe that for digit instance 5-8, the network is unable to produce a consistent prediction. In contrast to a single forward pass that might give a relatively large softmax output, computing the variance over the output of several forward passes will in this case indicate the high level of uncertainty in the network's prediction.

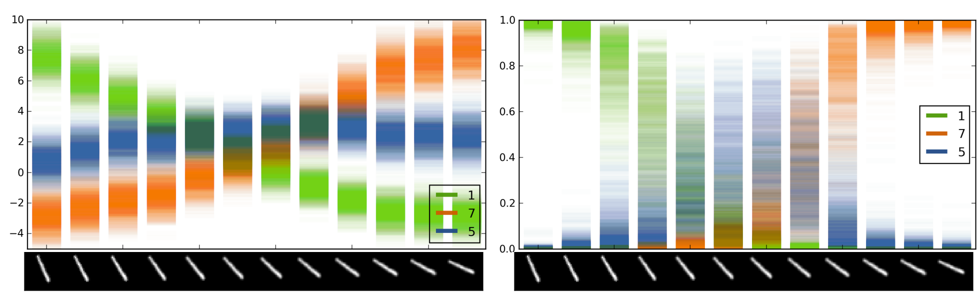

Figure 2: Softmax Input and Output for 100 forward passes of a rotated MNIST digit. The mean and variance of the Softmax output correspond to the prediction with uncertainty estimate obtained by MC Dropout.

References

This summary is solely based on my understanding of the original paper. All images used here are taken from the original paper as well. The paper can be found under the following link:
`https://arxiv.org/pdf/1506.02142.pdf`