Computational Statistics and Data Analysis (MVComp2)

Exercise 9

Lecturer Tristan Bereau

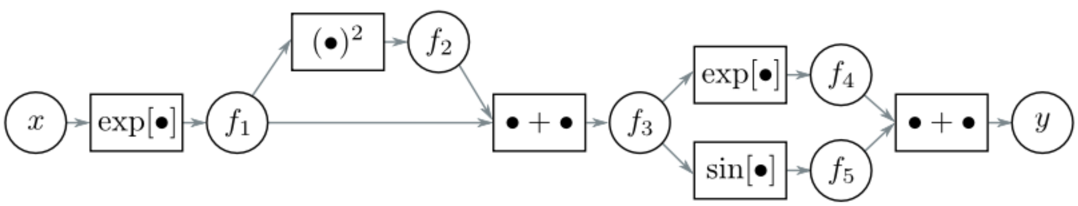
Semester Wi23/24
Due Jan. 11, 2024, 23:59

1 Forward- and reverse-mode differentiation (2 points)

We explore the computation of derivatives on general acyclic computational graphs. Consider the function

$$y = \exp\left[\exp(x) + \exp(x)^2\right] + \sin\left[\exp(x) + \exp(x)^2\right]$$

whose computational graph is depicted in the figure, together with the intermediate functions f_1, f_2, f_3, f_4, f_5.



(a) Compute the derivative $\partial y/\partial x$ by *forward-mode differentiation*. In other words, compute in order

$$\frac{\partial f_1}{\partial x}, \frac{\partial f_2}{\partial x}, \frac{\partial f_3}{\partial x}, \frac{\partial f_4}{\partial x}, \frac{\partial f_5}{\partial x}, \text{ and } \frac{\partial y}{\partial x},$$

using the chain-rule in each case to make use of the derivatives already computed.

(b) Compute the derivative $\partial y/\partial x$ by *reverse-mode differentiation*. In other words, compute in order

$$\frac{\partial y}{\partial f_5}, \frac{\partial y}{\partial f_4}, \frac{\partial y}{\partial f_3}, \frac{\partial y}{\partial f_2}, \frac{\partial y}{\partial f_1}, \text{ and } \frac{\partial y}{\partial x},$$

using the chain-rule in each case to make use of the derivatives already computed.

2 Second moment of ReLU (2 points)

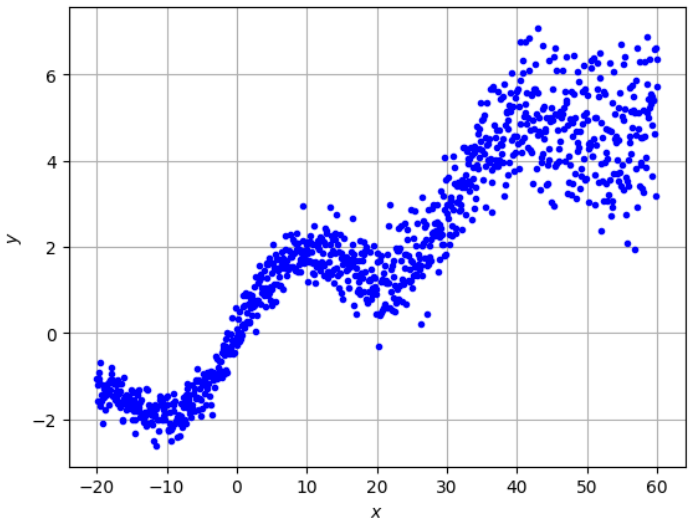
Consider the continuous random variable X with symmetrical distribution around the mean $E[X] = 0$ and variance $\mathrm{Var}[X] = \sigma^2$. e pass this variable through the ReLU function to obtain the transformed variable, B, such that

$$B(x) = \mathrm{ReLU}[x] = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}.$$

Prove that the second moment around the origin of the transformed variable is $E[B^2] = \sigma^2/2$.

3 Heteroskedastic regression (6 points)

Please download the following dataset: x_y.csv. It contains 1,000 datapoints with 1-dimensional inputs, $\boldsymbol{x}$, and 1-dimensional outputs, y, as shown in the following figure



(0) Download the dataset, and randomly split it into a training and a test set with ratio (66/33%).

(a) We model the distribution of outputs as

$$p(y|\boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}(y|f_\mu(\boldsymbol{x}), \sigma^2),$$

where $f_\mu(\boldsymbol{x})$ will be a multilayer perceptron (MLP), and we assume that all measurement errors are identical (i.e., homoskedastic regression), $\sigma(\boldsymbol{x}) = \sigma$. Explain why the mean-squared error is a reasonable log-likelihood function.

(b) Build an MLP with the following architecture: 3 fully-connected linear layers with hidden dimensions 32 and 64 (i.e., the dimensionality of your network should yield $1 \to 32 \to 64 \to 1$). Connect them with the ReLU activation function. Optimize the log-likelihood on the training set. Feel free to use any optimizer (e.g., stochastic gradient descent or Adam). Plot the resulting model on the test set together with the reference datapoints. Are you underfitting, overfitting?

(c) One standard way to address the prediction of heteroskedastic data for regression is to predict both the mean and the variance of a Normal distribution: $f_\mu(\boldsymbol{x}) = E[y|\boldsymbol{x}, \boldsymbol{\theta}]$ and $f_{\sigma^2}(\boldsymbol{x}) = \mathrm{Var}[y|\boldsymbol{x}, \boldsymbol{\theta}]$. Derive a log-likelihood function for this heteroskedastic problem.

(d) Modify the MLP to output two values: the mean and the variance. Optimize the model. Plot the mean and variance predicted by your model on the test set together with the reference datapoints. Are you underfitting, overfitting?

Hint: In case you use PyTorch, you may find the following functions useful to reshape tensors and arrays: `numpy.reshape` and `torch.squeeze`.