Exercise Sheet 3 - Gradient Descent

Prof. Dr. Paul Swoboda

NOTE: For this exercise sheet, you must submit a jupyter notebook and a PDF.

Exercise 1: Deriving the gradient (1 Points)

Assume we have a feature tensor $X \in \mathbb{R}^{N \times M}$ and multi-class labels of $\hat{Y} \in \mathbb{N}^N$ with each element $\hat{y}$ being $\hat{y} \in \{1, 2, ..., C\}$. To include our bias term we transform the feature tensor to have rows of 1 in the first column, i.e. $X \in \mathbb{R}^{N \times (M+1)}$. We now define our learnable parameters as $W \in \mathbb{R}^{(M+1) \times C}$. Derive the gradient ∇_w of the cross entropy function given by:

$$\mathcal{L} = -\sum_{i=1}^{N} \sum_{j=1}^{C} 1\left\{\hat{Y}_i = j\right\} \left[(X \cdot W)_{ij} - ln\left(\sum_{k=1}^{C} e^{(X \cdot W)_{ik}} \right) \right]$$

Exercise 2: (4 Points)

Download the jupyter notebook provided on Ilias. Implement gradient descent (1 Point), stochastic gradient descent (1 Point), and Netwon's method (2 Point) for multi-class classification using the MNIST dataset. Use the model and loss function specified in exercise 1.