Time	Group	Submission in Moodle; Mails with subject: [SMD2023]
Th. 12:00–13:00	A	lukas.beiske@udo.edu and tristan.gradetzke@udo.edu
Fr. 08:45–09:45	B	jonas.hackfeld@ruhr-uni-bochum.de and ludwig.neste@udo.edu
Fr. 10:00–11:00	C	stefan.froese@udo.edu and vincent.latko@udo.edu

Exercise 22 *Deep Learning Short Questions* **4 p.**

(a) What does the loss function describe and what is it used for?

(b) How can the loss function be minimised?

(c) What is the purpose of the activation functions and what problem do they solve? Name three common activation functions.

(d) What is a neuron?

(e) Give three examples of applications for neural networks and briefly describe why they are particularly suitable for these cases.

Exercise 23 *Linear Classification using Softmax* **6 p.**

For parameter tweaking in classification using the softmax function, the gradient of the cost function must be determined for all parameters to be adjusted. The cost function C is known from the lecture:

$$C(f) = \frac{1}{m}\sum_{i=1}^{m}\hat{C}(f_i) = \frac{1}{m}\sum_{i=1}^{m}\left[-\sum_{k=1}^{K}\mathbf{1}(y_i = k)\log\frac{\exp(f_{k,i})}{\sum_j \exp(f_{j,i})}\right]. \tag{1}$$

The chain rule is used to derive the loss function:

$$\nabla_W \hat{C} = \sum_{k=1}^{K}\frac{\partial \hat{C}}{\partial f_{k,i}}\cdot\frac{\partial f_{k,i}}{\partial W} \tag{2}$$

$$\nabla_b \hat{C} = \sum_{k=1}^{K}\frac{\partial \hat{C}}{\partial f_{k,i}}\cdot\frac{\partial f_{k,i}}{\partial b}. \tag{3}$$

(a) Given K classes and m examples x_i each with M components. What is the dimension of the individual components x_i, C, W, b, $\nabla_W\hat{C}$, $\nabla_{f_i}\hat{C}$, $\frac{\partial f_{k,i}}{\partial W}$, $\frac{\partial f_{k,i}}{\partial b}$? Also distinguish between row and column vectors.

(b) Show that the derivative of the loss function with respect to the scores for class a gives the following:

$$\nabla_{f_a} C(f) = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{\exp(f_{a,i})}{\sum_j \exp(f_{j,i})} - \mathbf{1}(y_i = a)\right]. \tag{4}$$

Here $\mathbf{1}(y_i = a)$ means that the entry in the unit matrix $\mathbf{1}$ is 1 only if y_i belongs to the class a. In all other cases the entries are 0.

(c) As a second step of the chain rule, show that the derivatives of $f_{k,i}$ with respect to W and b with $f_{k,i} = W_k x_i + b_k$ result in

$$\frac{\mathrm{d}f_{k,i}}{\mathrm{d}W} = \begin{array}{cl} \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \\ x_{i,1} & \dots & x_{i,M} \\ 0 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{pmatrix} & \begin{array}{l} \\ \\ \\ \leftarrow \text{row } k \\ \\ \\ \\ \end{array} \end{array}$$

and

$$\frac{\mathrm{d}f_{k,i}}{\mathrm{d}b} = \begin{array}{cl} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{array}{l} \\ \\ \\ \leftarrow \text{row } k \\ \\ \\ \\ \end{array} \end{array}$$

.

(d) Implement the linear classification with Softmax for the two populations `P_0` and `P_1` from the file `populations.hdf5` found in the moodle. Proceed as follows:

- Read in the populations with the keys `P_0` and `P_1`.
- Merge both populations and create the corresponding labels (`P1` has label 1).
- Initialise the weight matrix and the bias vector.
- Use a learning-rate of 0.5 and train 100 epochs.
- Implement the following steps vectorised (use `np.matmul`).
  - Implement the softmax function and the indicator function.
  - Iterate over the number of epochs. Calculate the softmax function for the current parameters W and b in each iteration. Using the values obtained, determine the gradient of the loss function with respect to W and b. Finally, update all parameters needed for the next iteration.

Hint: The gradient to W can also be calculated according to $\nabla_W C = \nabla_f C \cdot x_i^T$.

Exercise 24 *Ethical Problems in Datasets and Machine Learning* **0 p.**

Please note: You are not supposed to hand in any solution to this excercise. This task is meant to help you educate yourself about ethical implications in data science and statistics and discuss those with your peers. Of course this task will not suffice in discussing this topic exhaustively, it is only meant as an entry-point and as a mean to make you aware of these problems. We encourage you to participate in a discussion during the next tutoring group meetings and to educate yourself further on this topic.

The Boston Housing Dataset is a dataset formerly popular in teaching and testing Machine Learning. Originally created by D. Harrison Jr. and D. L. Rubinfeld for their publication “Hedonic housing prices and the demand for clean air"[1], the dataset was meant to create a model "to measure the willingness to pay for clean air". The approach chosen is called Hedonic Regression. The basic idea is to factor a good (e.g. a house or a car) into intrinsic and extrinsic characteristics (features) to estimate a dependent variable, in the present case the "median value of owner-occupied homes". For this, the authors collected data from multiple sources for the Boston, MA metropolitan area.

a) Educate yourself about the dataset. You can load the dataset via `scikit-learn`'s `load_boston`-function. Take a look at the features present, you can refer to the original publication or the `sklearn` documentation for the features meaning. Discuss why some features are problematic.

b) Regardless of the original study's target to study the impact of air-pollution on housing prices: Discuss, wether a model trained on this data is suitable to objectivly predict the value of a house.

c) Assume that a model derived from this data is used by e.g. real estate companies to value houses or a bank to decide whether a person can get a loan. Discuss the implications for society and especially underprivileged groups.

This is of course only a small subset of all the ethical considerations that occure in the context of statistics and machine learning. We encourage you to participate in the seminar "Ethik der Naturwissenschaften" (ethics of natural science – previously held in german language) in the winter term where you will have the time to discuss further problems.

[1] D. Harrison Jr. and D. L. Rubinfeld. "Hedonic housing prices and the demand for clean air." In: J. Environ. Econ. Manag. 5.1, 1978, pages 81-102