Machine Learning Exercise (SS 22)

Assignment 4: Logistic Regression

Mojtaba Nayyeri
mojtaba.nayyeri@ipvs.uni-stuttgart.de

Alexandra Baier
Alexandra.Baier@ipvs.uni-stuttgart.de

Akram Hosseini
Akram.Hosseini@ipvs.uni-stuttgart.de

Tim Schneider
timphillip.schneider@ipvs.uni-stuttgart.de

This assignment sheet consists of 4 pages with the theoretical tasks.

Submit your solution in ILIAS as a single PDF file.[1] Make sure to list full names of all participants, matriculation number, study program and B.Sc. or M.Sc on the first page. Optionally, you can *additionally* upload source files (e.g. PPTX files). If you have any questions, feel free to ask them in the exercise forum in ILIAS.

Submission is open until Monday, 23 May 2022.

[1] Your drawing software probably allows to export as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like pdfarranger) to combine the PDFs into a single file.

Classification with Linear Regression

Consider the following 2-dimensional input $x = \begin{bmatrix} 1 & -2 & 0.3 & 5 & 3 & 7 \\ 3 & 2 & 1 & -1 & 4 & 3 \end{bmatrix}$ with corresponding binary class labels y= [1;0;0;1;1;1]. Use (least-squares) linear regression, as shown in the lecture, to train on these samples and classify them. Your model should include an intercept term.

1. **Task**: Provide the coefficients of the linear regression (on x and y) and explain shortly how you computed them.
2. **Task**: Classify each of the 6 samples with your linear regression model. Explain how you map the continuous output of the linear model to a class label.
3. **Task**: Discuss in your own words, why linear regression is not suitable for classification.

Log-likelihood gradient and Hessian

Consider a binary classification problem with data $D = \{(x_i, y_i)\}_{i=1}^n$, $x_i \in \mathbb{R}^d$ and $y_i \in \{0, 1\}$. We define

$$f(x) = \phi(x)^T \beta \ , \quad p(x) = \sigma(f(x)) \ , \quad \sigma(z) = 1/(1 + e^{-z})$$

$$L^{\text{nll}}(\beta) = -\sum_{i=1}^{n} \Big[y_i \log p(x_i) + (1 - y_i) \log[1 - p(x_i)] \Big]$$

where $\beta \in \mathbb{R}^d$ is a vector. (Note: $p(x)$ is a short-hand for $p(y = 1|x)$.)

1. **Task**: Compute the derivative $\frac{\partial}{\partial \beta} L(\beta)$. Tip: Use the fact that $\frac{\partial}{\partial z}\sigma(z) = \sigma(z)(1 - \sigma(z))$.
2. **Task**: Compute the 2nd derivative $\frac{\partial^2}{\partial \beta^2} L(\beta)$.

Discriminative Function in Logistic Regression

Task: Logistic Regression defines class probabilities as proportional to the exponential of a discriminative function:

$$P(y|x) = \frac{\exp f(x, y)}{\sum_{y'} \exp f(x, y')}$$

Prove that, in the binary classification case, you can assume $f(x, 0) = 0$ without loss of generality.

This results in

$$P(y = 1|x) = \frac{\exp f(x, 1)}{1 + \exp f(x, 1)} = \sigma(f(x, 1)).$$

(Hint: First assume $f(x, y) = \phi(x, y)^T \beta$, and then define a new discriminative function f' as a function of the old one, such that $f'(x, 0) = 0$ and for which $P(y|x)$ maintains the same expressibility.)