University of Stuttgart | ANALYTIC COMPUTING

Machine Learning (SS 24)

Assignment 5: Logistic Regression

Yi Wang
yi.wang@ki.uni-stuttgart.de

Akram Sadat Hosseini
Akram.Hosseini@ki.uni-stuttgart.de

Rodrigo Lopez Portillo Alcocer
rodrigo.lopez-portillo-alcocer@ki.uni-stuttgart.de

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

Osama Mohammed
osama.mohammed@ki.uni-stuttgart.de

Daniel Frank
daniel.frank@ki.uni-stuttgart.de

Farane Jalali-Farahani
farane.jalali-farahani@ki.uni-stuttgart.de

Make sure to list the 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, 20th of May 2024, 12:00 noon.

Task 1: True or False: Justify Your Answers (15 Points)

1. In logistic regression, increasing the regularization parameter lambda (λ) decreases the magnitude of the model coefficients, potentially leading to underfitting if λ is set too high.
2. The output of a logistic regression model is a probability that the given input point belongs to the positive class.
3. Logistic regression models can be fitted using the least squares method as effectively as using maximum likelihood estimation.
4. The logistic function used in logistic regression can return values greater than 1.
5. Decision boundaries created by logistic regression are always linear.

Task 2: From Linear Regression to Classification (25 Points)

1. Consider the following 2-dimensional input $\mathbf{x}$:

$$\mathbf{x} = \begin{bmatrix} 1 & -2 & 0.3 & 5 & 3 & 7 \\ 3 & 2 & 1 & -1 & 4 & 3 \end{bmatrix}$$

with corresponding binary class labels $\mathbf{y} = [1; 0; 0; 1; 1; 1]$. 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.

   (a) Provide the coefficients of the linear regression (on x and y) and explain shortly how you computed them.

   (b) 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.

2. Research (e.g. see Hastie et al. chapter 4.1) and then explain what a discriminative function is and how it can be used for classification problems using the tools from linear regression.

Task 3: Log-likelihood gradient and Hessian (10 Points)

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

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

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

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

Note: The gradient and Hessian are needed to compute the optimal parameters for *logistic regerssion* models. Details on how to do this will be covered in the upcoming lecture.

1. 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. Compute the 2nd derivative $\frac{\partial^2}{\partial \beta^2} L(\beta)$.