Exercise for Machine Learning (SS 20)

Assignment 4: Logistic Regression

Prof. Dr. Steffen Staab, steffen.staab@ipvs.uni-stuttgart.de

Alex Baier, alex.baier@ipvs.uni-stuttgart.de

Janik Hager, janik-manel.hager@ipvs.uni-stuttgart.de

Ramin Hedeshy, ramin.hedeshy@ipvs.uni-stuttgart.de

Analytic Computing, IPVS, University of Stuttgart

Submit your solution in Ilias as either PDF for theory assignments or Jupyter notebook for practical assignments.
Mention the names of all group members and their immatriculation numbers in the file.
Submission is possible until the following Monday, 25.05.2020, at 14:00.

1 Classification with Linear Regression

Consider the following 1-dimensional input $x = [-2.0, -1.0, 0.5, 0.6, 5.0, 7.0]$ with corresponding binary class labels $y = [0, 0, 1, 0, 1, 1]$. 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. Provide the coefficients β of the linear regression (on x and y) and explain shortly how you computed them.
2. 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. Discuss in your own words, why linear regression is not suitable for classification.

2 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

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

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

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

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 $\frac{\partial^2}{\partial \beta^2} L(\beta)$.

3 Discriminative Function in Logistic Regression

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 $f(x,y) = \phi(x,y)^\top \beta$, 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.)