Machine Learning

Exercise Sheet 3

Winter Term 2023/2024
Prof. Dr. Niels Landwehr
Dr. Ujjwal

Available: 16.11.2023
Hand in until: 23.11.2023 11:59am
Exercise sessions: 27.11.2023/29.11.2023

Task 1 – Logistic Regression [25 points]

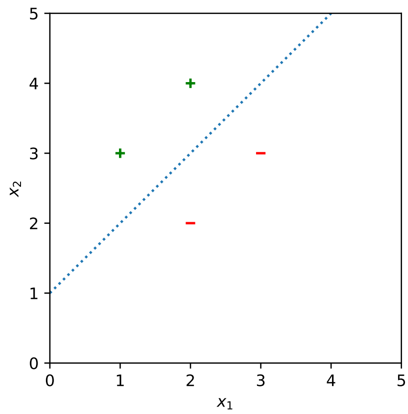
In this exercise, we study a binary logistic regression model with two-dimensional input of the form

$$\begin{aligned} f_{\boldsymbol{\theta}}(\mathbf{x}) &= \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2) && (1)\\ &= \sigma(\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2) && (2)\\ &= \sigma(\mathbf{x}^\mathsf{T}\boldsymbol{\theta}) && (3) \end{aligned}$$

where $x_1, x_2 \in \mathbb{R}$ are the two input features and as discussed in the lecture on linear regression we use an additional constant "dummy" feature $x_0 = 1$ to conveniently include the offset term θ_0 in the model. The model outputs the probability for the positive class, that is, $f_{\boldsymbol{\theta}}(\mathbf{x}) = p(y = 1|\mathbf{x}, \boldsymbol{\theta})$.
Consider the following set of $N = 4$ data points for this model given in matrix form:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 1 & 2 & 2 \\ 1 & 3 & 3 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad (4)$$

where the rows of $\mathbf{X}$ contain the instances $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ and $\mathbf{y}$ contains the targets y_1, y_2, y_3, y_4. We can visualize the data (ignoring the dummy attribute x_0) in a two-dimensional plot as follows:



In the figure, we have already included a possible linear decision boundary that correctly separates these four instances. The proposed decision boundary is characterized by running through the points $x_1 = 0$, $x_2 = 1$ and $x_1 = 4$, $x_2 = 5$.

a) Determine model parameters $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*, \theta_3^*)$ for the logistic regression model given by Equation 1 such that the model has the decision boundary proposed in the figure. Confirm that your model parameters correctly classify all instances. Compute the logarithmic likelihood of the data $\mathbf{X}, \mathbf{y}$ given this model.

b) Characterize the space of all model parameters $\boldsymbol{\theta}$ that would result in the same decision boundary and predictions as given by your model $\boldsymbol{\theta}^*$. That is, characterize the space

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta} | \forall \mathbf{x} \in \mathcal{X} : sign(\mathbf{x}^\top \boldsymbol{\theta}) = sign(\mathbf{x}^\top \boldsymbol{\theta}^*)\}$$

where

$$sign(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z = 0 \\ -1 & \text{if } z < 0 \end{cases} . \tag{5}$$

From the space $\boldsymbol{\Theta}$, find model parameters $\boldsymbol{\theta}^*_+$ that have a higher logarithmic likelihood than the original model $\boldsymbol{\theta}^*$, and parameters $\boldsymbol{\theta}^*_-$ that have a lower logarithmic likelihood.

Task 2 – Binary and Multiclass Logistic Regression [15 points]

Assume a multiclass logistic regression model for $T = 2$ classes of the form

$$f_{\mathbf{B}}(\mathbf{x}) = softmax(\mathbf{B}\mathbf{x}) \tag{6}$$

where $\mathbf{B} \in \mathbb{R}^{2 \times M}$ and we have left out the offset terms $\mathbf{b}$ because an offset can easily be incorporated by using a "dummy" feature that is constantly one as in Task 1. As explained in the lecture, the output of the model is a vector $f_{\mathbf{B}}(\mathbf{x}) \in \mathbb{R}^2$ of probabilities for the two classes.

a) From the matrix $\mathbf{B}$, construct a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^M$ for a binary logistic regression model such that the probability $p(y = 1|\mathbf{x}, \boldsymbol{\theta}) = \sigma(\mathbf{x}^\top \boldsymbol{\theta})$ obtained from the binary logistic regression is identical to the probability $p(y = 1|\mathbf{x}, \boldsymbol{\theta})$ obtained as the second element of the output vector $f_{\mathbf{B}}(\mathbf{x})$ of the multiclass logistic regression model. Here, we are assuming for the multiclass model that the second class is the positive class.

b) Construct a matrix $\mathbf{B} \in \mathbb{R}^{2 \times 3}$ such that for any input the class membership probabilities returned by the multiclass logistic regression model are identical to the class membership probabilities returned by the model $\boldsymbol{\theta}^*$ from Task 1 a).