W3WI DS304.1 Applied Machine Learning Fundamentals

Exercise Sheet # 4 - Logistic Regression

Question 1 (Why you should not use linear regression for classification)

The following dataset contains examples of **planets** and **dwarf planets**. Each celestial body in the dataset is described only by its radius (distance from core to surface) in millions of meters. Based on this feature we want to learn a classification model which predicts if the object is either a planet or a dwarf planet.

Table 1 and figure 1 introduce the dataset which contains $n = 6$ training instances:

Row	**Object**	**Radius** ($\times 10^6$ m)	**Label**	**Label encoded**
1	Ceres	1.0	dwarf planet	0
2	Eris	2.3	dwarf planet	0
3	Pluto	2.4	dwarf planet	0
4	Mercury	4.9	planet	1
5	Earth	12.8	planet	1
6	Jupiter	143.0	planet	1

Table 1: Dataset of planets and dwarf planets.

Figure 1: Visualization of the planet dataset (not drawn to scale).

Please answer the following questions:

1. As a baseline you decide to use a linear regression model. Compute the **decision boundary** using the normal equation. The last column in table 1 contains the encoded target labels. Apply the threshold $\rho := 0.5$, i.e. predict the positive class (planet), if the model output is greater or equal to 0.5, and the negative class (dwarf planet) otherwise. What problem do you observe?

2. Train a logistic regression classifier on the training data. To achieve this, implement the **batch gradient descent** algorithm for logistic regression (e.g. in Python). Use the learning rate $\alpha := 0.075$ and initialize the algorithm at the point $\boldsymbol{\theta}_0 := \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

   - What are the model parameters after 20 iterations of gradient descent?

   - Feed the training data into your trained classifier. What are the predictions (again use the threshold $\rho := 0.5$)?

- Compute the decision boundary generated by your model! *(Do this by hand!)* How does logistic regression perform compared to linear regression?
- Is your logistic regression model Bayes optimal?

Question 2 (Derivative of the sigmoid function)

The derivative of the sigmoid function

$$g(z) := \frac{1}{1 + \exp(-z)}$$

is crucial for the training of a logistic regression classifier. In the lecture slides we have seen that the derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}z} g(z) = g(z) \cdot \big(1 - g(z)\big).$$

Proof that this is the correct derivative of the sigmoid function! **Hint:** The quotient rule of differentiation might be helpful.

Question 3 (Stochastic gradient descent)

Let the training example $\boldsymbol{x} := (-3, 1, -1, 1, 0)^\intercal, y = 1$ be given. Perform one iteration of (stochastic) gradient descent starting with $\boldsymbol{\theta}_0 := (1, 2, 3, 4, 5)^\intercal$. Use the learning rate $\alpha := 0.2$. What are advantages and disadvantages of stochastic gradient descent? What other types of gradient descent do you know? Briefly explain the concepts.

Question 4 (Logistic regression with basis functions)

Write down the model function $h_{\boldsymbol{\theta}}\big(\boldsymbol{\varphi}(\boldsymbol{x})\big)$ of the model depicted in figure 2.

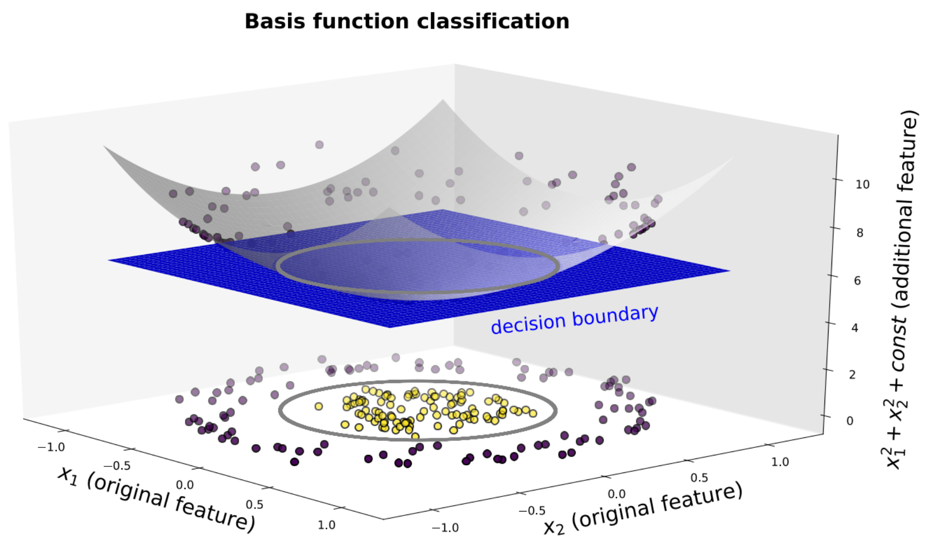

Figure 2: Logistic regression with polynomial basis functions.

Question 5 EX 2022 (Multi-class classification with logistic regression)

You want to automatically detect handwritten digits (10 possible classes). You decide to train a logistic regression classifier. You choose the **One-vs-One** approach since logistic regression is a binary classification model.

How many (binary) classifiers do you have to train to complete the task?

Question 6 (Multi-class classification with logistic regression)

Compare the number of base classifiers to be trained in One-vs-One and One-vs-Rest for different values of K (number of classes). Which method is more expensive for large K?