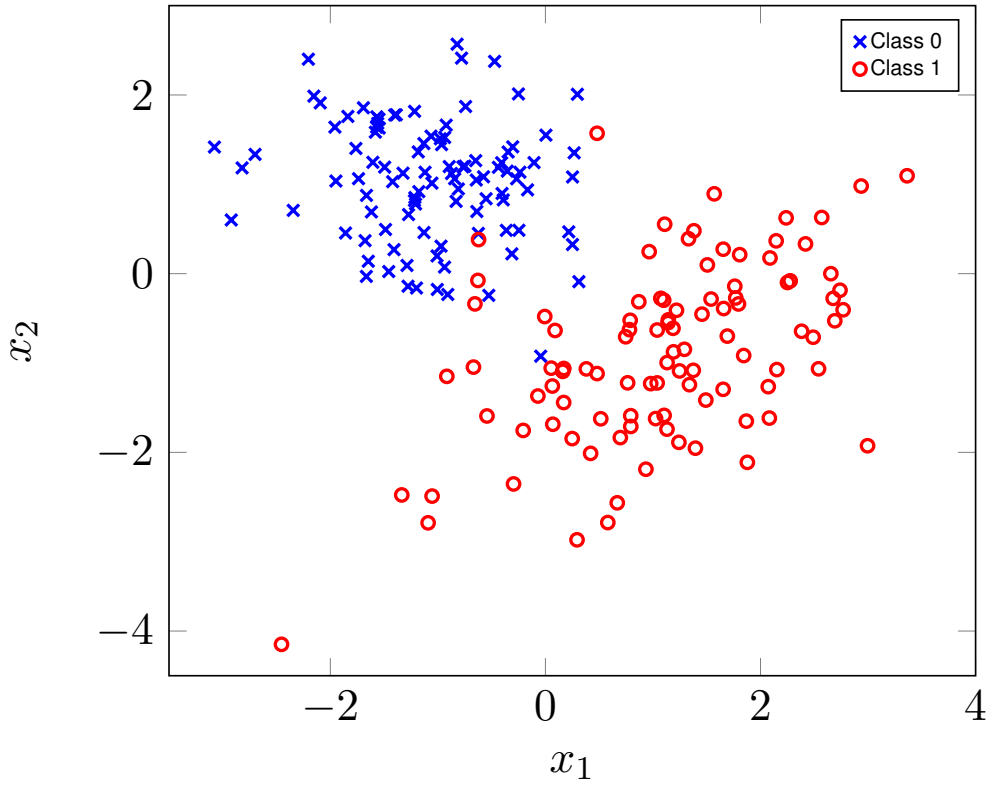
Chair of Decision Sciences and Systems
TUM School of Computation, Information and Technology
Technical University of Munich

Business Analytics & Machine Learning
Tutorial sheet 3: Logistic Regression

Prof. Dr. Martin Bichler, Prof. Dr. Jalal Etesami
Julius Durmann, Markus Ewert, Johannes Knörr, Yutong Chao

Exercise T3.1 *Logistic regression for a 2D classification problem*

You are given the data set in *2d-classification-data.csv* which is also visualized below. The data consists of two-dimensional points belonging to one of two classes: 0 or 1.



In this exercise, we are going to find a predicting model which can classify new data points. Consider the logistic regression model:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\Longleftrightarrow$$

$$p(y=1|x_1,x_2) = \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2}}$$

a) For which values $x = (x_1, x_2)$ does the logistic regression model output $p(y = 1|x) = p(y = 0|x)$? Derive a functional description $x_2 = f(x_1)$ which describes the corresponding decision boundary.

b) Without computation, draw a straight line which roughly separates the two classes. Consider this as a decision rule whether a sample belongs to class 0 or class 1: How many samples from the data set are miss-classified (attributed to the wrong class)?

c) Using Python and scikit-learn (or statsmodels), derive the optimal parameters for the logistic regression model. If you like, you may use the provided template notebook for this purpose.

d) Draw the decision boundary of the optimal model. How many samples are miss-classified?
Can you explain why the model might perform worse (w.r.t the number of miss-classified samples) than your own decision boundary from b)?

Exercise T3.2 *Maximum likelihood estimation*

You are given the following dataset with the dependent binary variable y and the independent variable x.

x	y
1	0
2	0
2.5	1
4	1

Based on these data points we want to create a logistic regression model with the logistic function σ (or more broadly a sigmoid function):

$$\text{Pr}[Y|X] = p(x) = \sigma(\beta_0 + \beta_1 x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}.$$

To estimate the logistic regression coefficients, we will use *maximum likelihood estimation*.
To simplify notation, let $p_i = p(x_i) = \sigma(z_i)$ and $z_i = \beta_0 + \beta_1 x_i$.

a) Determine the likelihood function $L(\beta)$.
Hint: To keep everything simple, it is sufficient to formulate L in terms of p_i (which includes the dependency on β).

b) Find the gradient for the log of the likelihood function $LL(\beta)$. The gradient is defined as:

$$\nabla LL(\beta) = \begin{pmatrix} \frac{\partial LL(\beta)}{\partial \beta_0} \\ \frac{\partial LL(\beta)}{\partial \beta_1} \end{pmatrix}.$$

Hint: Use the chain rule: $\frac{\partial LL}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial LL}{\partial p_i} \frac{\partial p_i}{\partial z_i} \frac{\partial z_i}{\partial \beta_j}$ with $z_i = \beta_0 + \beta_1 x_i$. You may use the following derivative of the logistic function σ without proof: $\sigma'(z_i) = \sigma(z_i)(1 - \sigma(z_i))$.

c) Given the initial values $\beta^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and a learning rate $\alpha = 0.2$, calculate the coefficients after the first iteration of gradient ascent.

d) If a linear regression model was fitted to a logistic regression dataset, what could be the problems w.r.t. the Gauss Markov properties?

Exercise T3.3 *Poisson regression*

You are provided the following numbers from the result of a *Poisson regression model*.

Variable	Estimate	Std. Error
Intercept	1.5499	0.0503
Age	-0.0047	0.0009

a) According to the model above, what *qualitative* effect does a change in the independent variable age (+1) have on the dependent variable dv.

b) According to the model above, what *quantitative* effect (on the incidence rate and log-incidence rate) does a change in the independent variable age (+1) have on the dependent variable dv.