Homework 2: Data normalization

Data are often normalized. Each feature subtracts its mean and then divides the result by its standard deviation. The normalized features have zero mean and unit standard deviation. This may help in several cases:

- When each feature has a different order of magnitude (such as millimetres and kilometres). Then the gradient would ignore the feature with the smaller values.

- When problems such as vanishing gradients are present (we will elaborate on this in Exercise 4).

Task:

Write function normalize which takes as an input a dataset and normalizes it. Then train the same classifier as in the lecture "logistic regression". Use the original and normalized dataset. Which differences did you observe when

- the logistic regression is optimized via gradient descent?

- the logistic regression is optimized via Newton's method?

Do you have any intuition as to why? Write a short report (in Latex) summarizing your findings.