Sheet #6 – Solution

Due:

Fred Hamprecht
Manuel Haußmann

1 Properties of the Logistic Sigmoid (9 pt)

The logistic sigmoid function is defined as

$$\sigma(a) = \frac{1}{1+\exp(-a)}.$$

Show that the logistic sigmoid satisfies the following properties given in the lecture:

i) $\sigma(-a) = 1 - \sigma(a)$,

ii) $\frac{d}{da}\sigma(a) = \big(1-\sigma(a)\big)\sigma(a)$,

iii) $\sigma^{-1}(b) = \log \frac{b}{1-b}$.

Solution:

i)

$$1-\sigma(a) = 1 - \frac{1}{1+\exp(-a)} = \frac{1+\exp(-a)-1}{1+\exp(-a)} = \frac{1}{1+\exp(a)} = \sigma(-a)$$

ii)

$$\frac{d}{da}\sigma(a) = \frac{d}{da}\frac{1}{1+\exp(-a)} = \frac{\exp(-a)}{(1+\exp(-a))^2} = \frac{\exp(-a)}{1+\exp(-a)}\frac{1}{1+\exp(-a)} = (1-\sigma(a))\sigma(a)$$

iii)

$$\begin{aligned} b = \sigma(a) = \frac{1}{1+\exp(-a)} \quad &\Leftrightarrow \quad b(1+\exp(-a)) = 1 \\ \Leftrightarrow \quad \exp(-a) = \frac{1-b}{b} \quad &\Leftrightarrow \quad a = \log\frac{b}{1-b} \end{aligned}$$

2 Comparison of Logistic to Linear Regression (9 pt + 5pt)

The goal of this exercise is to reproduce the 1d example discussed in the lecture, comparing your existing linear regression implementation to the proposed logistic regression.

i) Implement *iteratively reweighted least-squares (IRLS)* and fit it to the `log-reg.npz` data set (starting from a random $\boldsymbol{\beta}$ estimate). Visualize the change in the data fit after each update step.

ii) Add an outlier point ($x = 10$) to the data and refit the model. Compare your original solution with this new model. What do you observe?

iii) Fit your OLS solution to both data sets and compare it with the results you got via the logistic regression approach.

iv) *(technical + 5pt)* Construct a similar 1d data set that is linearly separable and observe the convergence behavior of your model. Extend the logistic regression loss with a regularization term as for the ridge regression and fit this regularized model. Does the behavior change?

Solution: See jupyter-notebook.

3 Fitting Logistic Regression to multidimensional data (15 pt)

The task of this exercise is to fit a logistic regression to multidimensional data.

i) Consider the data set `log-reg2.npz`. It is a binary problem where each class consists of a noisy circle and one circle lies within the second. Plot the data, coloring the points by the class label.

ii) In order to fit this nonlinear data, but still be able to plot the decisions, extend the spatial data $\mathbf{X}$ to $\tilde{\mathbf{X}}$ which consists of an extended set of features using this function.

```
def create_features(X, degree=5):
  data = np.stack([X[0]**i * X[1]**(d-i) for d in range(0,degree) for i in range(d+1)])
  return data
```

Implement it and explain what kind of features it creates.

iii) Extend your logistic regression implementation from exercise 2 to be able to handle multidimensional features and to allow for ridge/L2 regularization, i.e. we are now optimizing

$$-\sum_{n=1}^{N} \log p(y_n|\boldsymbol{\beta}, \mathbf{x}_n) + \lambda||\boldsymbol{\beta}||_2^2.$$

Fit a logistic regression model to the data starting from $\boldsymbol{\beta} = \mathbf{0}$.

Hint: For this exercise, it is sufficient to directly work with the regularized gradient and Hessian from the lecture without first deriving a new IRLS like update rule.

iv) Plot your predictions for a grid of test points on the interval $[-1.5, 1.5] \times [-1.5, 1.5]$ for several λ and discuss your observations on changes in the decision boundary you might observe.

Solution: See jupyter-notebook.