Exercise 3 – Classification I

Introduction to Machine Learning

Exercise 1: Logistic vs softmax regression

This exercise is only for lecture group A

Learning goals

Solve "show equivalence"-type questions

Binary logistic regression is a special case of multiclass logistic, or softmax, regression. The softmax function is the multiclass analogue to the logistic function, transforming scores $\theta^\top \mathbf{x}$ to values in the range [0, 1] that sum to one. The softmax function is defined as:

$$\pi_k(\mathbf{x}|\theta) = \frac{\exp(\theta_k^\top \mathbf{x})}{\sum_{j=1}^{g} \exp(\theta_j^\top \mathbf{x})}, k \in \{1, ..., g\}$$

Show that logistic and softmax regression are equivalent for $g = 2$.

Exercise 2: Hyperplanes

Learning goals

1. Understand that hyperplanes bisect the space with a linear boundary
2. Get a feeling for coefficients in hyperplane equations

Linear classifiers like logistic regression learn a decision boundary that takes the form of a (linear) hyperplane. Hyperplanes are defined by equations $\theta^\top \mathbf{x} = b$ with coefficients θ and a scalar $b \in \mathbb{R}$.

In order to see that such expressions actually describe hyperplanes, consider $\theta^\top \mathbf{x} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0$. Sketch the hyperplanes given by the following coefficients and explain the difference between the parameterizations:

- $\theta_0 = 0, \theta_1 = \theta_2 = 1$
- $\theta_0 = 1, \theta_1 = \theta_2 = 1$
- $\theta_0 = 0, \theta_1 = 1, \theta_2 = 2$

Exercise 3: Decision Boundaries & Thresholds in Logisitc Regression

Learning goals

1) Understand that logistic regression finds a linear decision boundary
2) Get a feeling for how parameterization changes predicted probabilities

In logistic regression (binary case), we estimate the probability $p(y = 1|\mathbf{x}, \theta) = \pi(\mathbf{x}|\theta)$. In order to decide about the class of an observation, we set $\hat{y} = 1$ iff $\pi(\mathbf{x}|\theta) \geq \alpha$ for some $\alpha \in (0, 1)$.

Show that the decision boundary of the logistic classifier is a (linear) hyperplane.

Hint

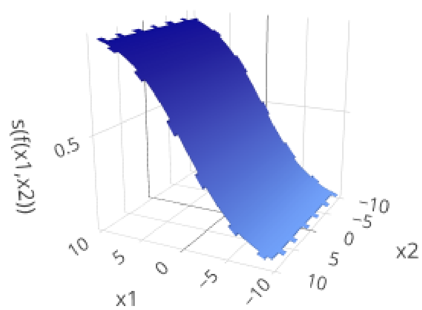
Derive the value of $\theta^\top \mathbf{x}$ (depending on α) starting from which you predict $\hat{y} = 1$ rather than $\hat{y} = 0$.

Below you see the logistic function for a binary classification problem with two input features for different values $\theta^\top = (\theta_1, \theta_2)^\top$ (plots 1-3) as well as α (plot 4). What can you deduce for the values of θ_1, θ_2, and α? What are the implications for classification in the different scenarios?

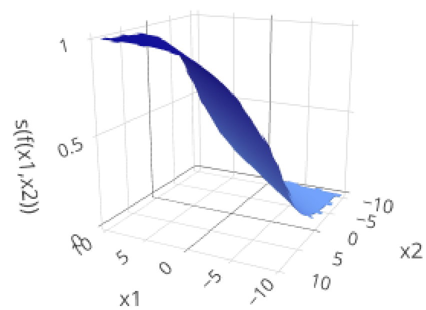
Derive the equation for the decision boundary hyperplane if we choose $\alpha = 0.5$.

Explain when it might be sensible to set α to 0.5.

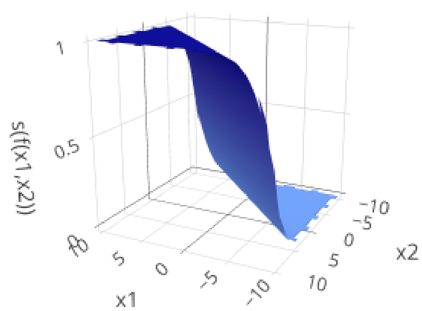
Plot (1)



Plot (2)



Plot (3)



Plot (4)