Exercise Sheet 2

Neural Computation
University of Birmingham

Week 2 2025

Exercise 1. Consider the the problem of finding an accurate temperature estimate $\hat{x}$ from a series of noisy measurements $\mathbf{x} = (x_1, \ldots, x_n)$. We want to find the estimate that minimises the sum of squared errors

$$x^* = \arg\min_{\hat{x}} L(\hat{x}, \mathbf{x}), \tag{1}$$

where

$$L(\hat{x}, \mathbf{x}) = \sum_{i=1}^{n} (\hat{x} - x_i)^2. \tag{2}$$

For this particular problem, there is an analytical solution for x^*, that is, a solution that we can find without requiring optimisation methods like gradient descent. Find this solution. Here are some helpful tips:

- Find the minimum of $L(\hat{x}, \mathbf{x})$ by setting its derivative $\frac{\partial}{\partial \hat{x}} L(\hat{x}, \mathbf{x}) = 0$ and solving for $\hat{x}$.
- The derivative of the sum is the sum of the derivatives of the summands.
- Remember that for a quadratic equation $(a - b)^2 = a^2 - 2ab + b^2$.

Exercise 2. Scientists interested in the penguin population on an Antarctic island counted the number of animals three times, resulting in population numbers of 332, 353, and 419, obtained in the years 2010, 2012, and 2016, respectively.

They want to fit a linear model (a perceptron without the sign function) to the data, describing the development of penguin numbers over time. To make the calculation easier, they express the years relative to the year of the first measurement. They initialise their model parameters by hand using a rough guess with slope $w^0 = 10$ and intercept $b^0 = 350$.

(a) According to this rough initial model, what would be the projected number of penguins in the year 2025?

(b) The scientists want to update their model using gradient descent, trying to minimise the sum of absolute errors

$$\sum_{i=1}^{n} |\hat{y}_i - y_i|$$

as their loss function. Compute the loss of this model with the initial parameters w^0, b^0.

(c) Compute the first gradient update for w and b, using a learning rate of $\eta = 0.5$.

Tip: The derivative of the absolute value function is not defined at zero, but for non-zero values it can be written as:

$$\frac{d}{dx}|x| = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$$

Use this to compute the gradient of the loss with respect to w and b.

(d) Compute the loss after the gradient step.

Exercise 3. There are two species of penguins living on the island: Adelie and Gentoo penguins. Consider a perceptron model for classifying the penguin species based on flipper length in cm (x_1) and body weight in kg (x_2). The model uses a logistic function as the activation function and outputs the probability that a penguin is a Gentoo penguin. Its parameters are $\mathbf{w} = (1, 1)$ and $b = -25$.

(a) Consider three penguins with flipper lengths of 22, 22, and 23 cm, and body masses of 2, 3, and 4 kg, respectively. Compute their probabilities of being Gentoo penguins according to the model.

(b) We want to change the weight vector to be $\mathbf{w}' = (2, 2)$ but adjust $b' = -50$. Compute the probabilities for the three penguins again using the new values. Does the new model have a different decision boundary? How can you describe the influence of the length of the weight vector on the model?