Machine Learning for Image Analysis, SoSe 2023
Exercise Sheet #2: Logistic Regression
Due date: May 2, 2023, before 11 am

Problem 1 (Properties of the Logistic Sigmoid) (9 pt).
A more common notation for the logistic sigmoid function uses σ and is defined as

$$\sigma(a) = \frac{1}{1 + e^{-a}} \tag{1.1}$$

Show that the logistic sigmoid satisfies the following properties:

(a) (4 pt) $\sigma(-a) = 1 - \sigma(a)$

(b) (5 pt) $\frac{d}{da}\sigma(a) = (1 - \sigma(a))\sigma(a)$

Problem 2 (Gradient of the Binary Cross-entropy Loss) (16 pt).
Your goal is to show step-by-step that the partial derivative of the logistic regression loss function (aka binary cross-entropy loss) is as follows

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i \tag{2.1}$$

Here x^i and θ are vectors in $\mathbb{R}^p$, $y^i \in \{0, 1\}$ and $J(\theta)$ and $h_\theta(x)$ are defined as in the lecture:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} -y^i \log(h_\theta(x^i)) - (1 - y^i) \log(1 - h_\theta(x^i)) \tag{2.2}$$

$$h_\theta(x) = g(\theta^T x) \tag{2.3}$$

$$g(z) = \frac{1}{1 + e^{-z}} \tag{2.4}$$

(a) (5 pt) Show that $J(\theta)$ can be formulated as

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} -y^i \theta^T x^i + \theta^T x^i + \log(1 + e^{-\theta^T x^i})$$

(b) (3 pt) Show that the term in (a) can be simplified to

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} -y^i \theta^T x^i + \log(1 + e^{\theta^T x^i})$$

(c) (3 pt) Show that

$$\frac{\partial}{\partial \theta_j} \theta^T x^i = x_j^i$$

(d) (5 pt) Using (b) and (c) show that

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) x_j^i$$

Hint: The rules for logarithms might be very helpful here.