1) Sigmoid function derivatives $\sigma(\eta)$

The sigmoid function is written as $\sigma(\eta) = \frac{1}{1+e^{-\eta}} = \frac{e^{\eta}}{1+e^{\eta}}$, where $0 < \sigma(\eta) < 1$.
Show that $\frac{d\sigma(\eta)}{d\eta} = \sigma(\eta)\,[1 - \sigma(\eta)]$ and $\frac{\mathrm{dlog}\sigma(\eta)}{d\eta} = 1 - \sigma(\eta)$.

2) Logistic Regression Likelihood & Cross-Entropy

Let $\mathcal{D} = \{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_N}, y_N)\}$, where $\mathbf{x_n} \in \mathbb{R}^D$ and $y_n \in \mathbb{R}$, be the training data of a binary logistic regression model with weights $\mathbf{w} \in \mathbb{R}^D$. The probability of sample $(\mathbf{x_n}, y_n)$ belonging to class 1 is $p(y = 1|\mathbf{x}, \mathbf{w}) = \sigma(\mathbf{w}^T\mathbf{x})$, while the probability of belonging to class 0 is $p(y = 0|\mathbf{x}, \mathbf{w}) = 1 - \sigma(\mathbf{w}^T\mathbf{x})$. Compute the likelihood $\mathcal{L}(\mathcal{D}|\mathbf{w})$ of data $\mathcal{D}$ given the model parameters $\mathbf{w}$, as well as the cross-entropy error $\mathcal{E}(\mathbf{w}) = -\mathrm{log}\mathcal{L}(\mathcal{D}|\mathbf{w})$.

3) Logistic Regression Optimization

3a) Show that the first order derivative (i.e., gradient vector) of the cross-entropy function is $\nabla\mathcal{E}(\mathbf{w}) = \frac{\vartheta\mathcal{E}(\mathbf{w})}{\vartheta\mathbf{w}} = \sum_{n=1}^{N} \underbrace{\left(\sigma(\mathbf{w}^T\mathbf{x_n}) - y_n\right)}_{\text{error}} \mathbf{x_n}$

3b) Show that the Hessian of the cross-entropy function is $\mathbf{H} = \frac{\vartheta^2\mathcal{E}(\mathbf{w})}{\vartheta^2\mathbf{w}} = \nabla\left((\nabla\mathcal{E}(\mathbf{w}))^T\right) = \sum_{n=1}^{N} \sigma(\mathbf{w}^T\mathbf{x_n}) \cdot \left(1 - \sigma(\mathbf{w}^T\mathbf{x_n})\right) \cdot \left(\mathbf{x_n} \cdot \mathbf{x_n}^T\right)$ and show that it is positive semi-definite.