Assignment 1. Prove that the family of univariate normal distributions $\mathcal{N}(\mu, \sigma)$ with density

$$p_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

is an exponential family

$$p_\eta(x) = \exp\left[\langle \phi(x), \eta \rangle - A(\eta) \right]$$

with sufficient statistic $\phi(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$. Deduce a formula expressing the natural parameter vector $\eta$ in terms of $\mu$ and $\sigma$.

Assignment 2. The probability density function of a Laplace distribution (aka double exponential distribution) with location parameter $\mu$ and scale $b$ is given by

$$p(x | \mu, b) = \frac{1}{b} \exp\left(-\frac{|x - \mu|}{b}\right).$$

Find the maximum likelihood estimates of the location parameter and the scale parameter given an i.i.d. sample $T_m = \{x_i \in \mathbb{R} | i = 1, \ldots, m\}$.

Assignment 3. Consider the family of multivariate normal distributions $\mathcal{N}(\mu, V)$ with density

$$p_{\mu, V}(x) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)^T V^{-1} (x - \mu)\right],$$

where $|V|$ denotes the determinant of the matrix $V$. It is parametrised by the mean $\mu \in \mathbb{R}^n$ and the symmetric covariance matrix $V$. Given i.i.d. training data $T_m = \{x^j \in \mathbb{R}^n | j = 1, \ldots, m\}$, we want to estimate the parameters $\mu$ and $V$ by MLE.

a) Show that the log-likelihood of the training data is given by

$$L(\mu, V, T^m) = \mathbb{E}_{T^m}\left[-\frac{1}{2}(x - \mu)^T V^{-1} (x - \mu)\right] + \frac{1}{2} \log |V^{-1}| + c.$$  

b) Compute the gradient w.r.t. $\mu$ and prove that the MLE estimate for it is given by $\mu^* = \mathbb{E}_{T^m}[x]$.

c) Compute the gradient w.r.t. $V^{-1}$. Use the fact that

$$\frac{\partial}{\partial A} \log |A| = A^{-T}$$

holds for any invertible symmetric matrix $A$. Prove that the MLE estimate for $V$ is given by

$$V^* = \mathbb{E}_{T^m}[(x - \mu^*)(x - \mu)^T].$$
Assignment 4. Consider the binary logistic regression model
\[ p(y \mid x) = \frac{e^{g(w, x)}}{2 \cosh \langle w, x \rangle}, \]
where \( x \in \mathbb{R}^n \) is a feature vector and \( y = \pm 1 \) is the object class. The model is parametrised by the vector \( w \in \mathbb{R}^n \). Given an i.i.d. training set \( \mathcal{T}^m = \{(x^j, y^j) \mid j = 1, \ldots, m\} \), we want to estimate the unknown parameter vector \( w \) of the model by maximising the conditional log-likelihood
\[ L(w, \mathcal{T}^m) = \mathbb{E}_{\mathcal{T}^m} \left[ \log(p(y \mid x)) \right] = \mathbb{E}_{\mathcal{T}^m} \left[ y \langle w, x \rangle - \log \cosh \langle w, x \rangle \right] - \log 2 \to \max_w \]
Prove that the objective function is concave in \( w \) by computing its second derivative (matrix) and showing that it is negative semi-definite.

Assignment 5. The Kullback-Leibler divergence for probability densities \( p(x) \) and \( q(x) \) is defined by
\[ D_{KL}(p \parallel q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, dx. \]
Compute the KL-divergence for two univariate normal distributions \( \mathcal{N}(\mu, \sigma) \) and \( \mathcal{N}(\tilde{\mu}, \tilde{\sigma}) \).