Assignment 1. Consider the following parameter estimation task. You are given i.i.d. training data \( T = \{ x_i \in \mathbb{R} \mid i = 1, 2, \ldots, m \} \) generated from the normal distribution
\[
p_{\mu_0}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2}}
\]
and the task is to estimate its unknown mean \( \mu_0 \).

a) Prove that the expected log-likelihood
\[
L(\mu) = \int_{-\infty}^{\infty} p_{\mu_0}(x) \log p_{\mu}(x) \, dx
\]
has a unique global maximum at \( \mu_0 \).

b) Show that the maximum likelihood estimator is given by the arithmetic mean of the training data, i.e.
\[
\mu^* = e_{ML}(T) = \frac{1}{m} \sum_{i=1}^{m} x_i,
\]
where
\[
e_{ML}(T) = \arg \max_{\mu} \frac{1}{m} \sum_{i=1}^{m} \log p_{\mu}(x).
\]
Prove that this estimator is unbiased.

c) Compute the variance of the maximum likelihood estimator, i.e.
\[
\mathbb{E}_{\mu_0}[ (\mu_0 - e_{ML}(T))^2 ].
\]
How does it depend on \( \mu_0 \) and \( m \)?

Assignment 2. Consider the exponential family,
\[
p_u(x) = \frac{1}{Z(u)} \exp \langle \phi(x), u \rangle
\]
where \( u \in \mathbb{R}^k \) is a parameter vector, \( \phi(x) \in \mathbb{R}^k \) is a feature map, \( x \in \mathcal{X} \) and
\[
Z(u) = \sum_x \exp \langle \phi(x), u \rangle
\]
is a normalising factor. This defines a parametrised class of probability distributions. (Show that the class of univariate normal distributions is an particular example of such a family.)

a) Prove that each model in this class is identifiable, provided that the affine hull of the set of vectors \( \{ \phi(x) \mid x \in \mathcal{X} \} \) is the entire space \( \mathbb{R}^k \) (or, equivalently, there is no hyperplane containing all vectors).
b) Derive the formula for the log-likelihood of given training data $T^m = \{(x^i) \mid i = 1, 2, \ldots, m\}$. Prove that the logarithm of the probability
\[
\log p_u(x) = \langle \phi(x), u \rangle - \log Z(u)
\]
is a concave function of $u$ by verifying the following steps.

1) Prove that the gradient of $\log Z(u)$ is
\[
\nabla_u \log Z(u) = \sum_x p_u(x) \phi(x) = \mathbb{E}_u(\phi).
\]

2) Prove that the second derivative of $\log Z(u)$ is
\[
\nabla_u^2 \log Z(u) = \sum_x p_u(x) \phi(x) \otimes \phi(x) - \mathbb{E}_u(\phi) \otimes \mathbb{E}_u(\phi) = \\
= \mathbb{E}_u[(\phi - \mathbb{E}_u(\phi)) \otimes (\phi - \mathbb{E}_u(\phi))]
\]

3) Deduce that the second derivative is a positive semi-definite matrix and conclude that $\log Z(u)$ is convex.

c) Suppose that the parameter vectors are bounded by $\|u\| \leq R$ and assume that the components of the vectors $\phi(x)$ are bounded in some interval $[a, b]$. Prove the Uniform Law of Large Numbers for the Maximum Likelihood Estimator by performing the following steps

1) Denote the log-likelihood of the training data $T^m$ by $L(u, T^m)$ and the expected log-likelihood by $L(u) = \mathbb{E}_v L(u, T^m)$, where $v \in \mathbb{R}^k$ is the true but unknown model.

2) Deduce that
\[
L(u, T^m) - L(u) = \langle \mathbb{E}_{T^m} \phi - \mathbb{E}_v(\phi), u \rangle
\]
holds, where $\mathbb{E}_{T^m} \phi$ denotes the arithmetic mean of the vectors $\phi(x)$ on the training data and $\mathbb{E}_v(\phi)$ denotes their expectation w.r.t. the true model.

3) Prove that
\[
\max_{\|u\| \leq R} |L(u, T^m) - L(u)| = \|\mathbb{E}_{T^m} \phi - \mathbb{E}_u(\phi)\| R
\]
holds.

4) Conclude the ULLN for MLE-s in this model class.