**Assignment 1** (ML with noisy labels). We want to learn a binary classifier \( q(k \mid x; \theta) \) with classes \( k = \pm 1 \). It is defined as a neural network with parameters \( \theta \) and with the sigmoid logistic distribution in the output.

The true labels \( k_i \) of the images \( x_i \) are however unknown. Instead we are given training pairs \((x_i, t_i)\) with “noisy labels” \( t_i = \pm 1 \). They might have been incorrectly assigned by the person who annotated the data. More specifically, let us assume that the label \( t_i \) is correct \((t_i = k_i)\) with probability \( 1 - \varepsilon \) and incorrect \((t_i = -k_i)\) with probability \( \varepsilon \).

a) Formulate the conditional maximum likelihood learning of the parameters \( \theta \).

**Hint:** the conditional likelihood of the training data sample \((x_i, t_i)\) is obtained by marginalizing over the unknown true label

\[
p(t_i \mid x_i) = \sum_{k \in \{-1, 1\}} p(t_i \mid k) q(k \mid x_i ; \theta),
\]

where \( p(t \mid k) \) is the labelling noise model.

b) A popular practical solution is to minimize the cross-entropy loss

\[
- \sum_i \sum_k p_i(k) \log q(k \mid x_i ; w), \tag{1}
\]

where \( p_i(k) \) denote "softened 1-hot labels": \( p_i(k) = 1 - \varepsilon \) for \( k = t_i \) and \( \varepsilon \) otherwise. Prove that the negative cross-entropy (1) is a lower bound of the log likelihood in a). Use Jensen’s inequality for \( \log \).

**Assignment 2.** Let \( q(x) \) and \( p(x) \) be two factorizing probability distributions for random vectors \( x \in \mathbb{R}^n \), i.e.

\[
p(x) = \prod_{i=1}^n p(x_i) \quad \text{and} \quad q(x) = \prod_{i=1}^n q(x_i).
\]

Prove that their KL-divergence decomposes into a sum of KL-divergences for the components, i.e.

\[
D_{KL}(q(x) \parallel p(x)) = \sum_{i=1}^n D_{KL}(q(x_i) \parallel p(x_i))
\]

**Assignment 3.** Compute the KL-divergence of two univariate normal distributions.
Assignment 4 (AP vs Triplet Loss). Starting with the expression for AP (see eq. (3) in the metric learning lab):

\[ AP = 1 - \frac{1}{T} \sum_{p \in P} \sum_{n \in N} \frac{[d_n < d_p]}{k(p)}, \]  

(2)

Verify that \( [z] \leq \max(z/\alpha + 1, 0) \) holds for each \( \alpha > 0 \) and use it as an approximation in the numerator of (2). How the resulting approximate AP is related to the triplet loss we used in the metric learning lab?