Machine Intelligence II **Exercise Sheet 1**

SS2012, Prof. Obermayer due: 26.04.2010

Problem 2.1 (5 points)
Let x be a random variable that can take the set of values $X = \{x_1, ..., x_n\}$. The Kullback-Leibler (KL) divergence is a measure of the difference between two probability distributions $p(x)$ and $q(x)$ and is defined as

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

a) (2 points)
A real-valued function f is called convex for any two points x and y and any $\lambda \in [0, 1]$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Jensen's inequality states that for a random variable x and a convex function $f(x)$

$$f(\mathbb{E}[x]) \leq \mathbb{E}[f(x)]$$

where $\mathbb{E}[.]$ denotes the expectation. Use this to show that the KL divergence is non-negative.

b) (3 points)
Show that the KL divergence is not symmetric by finding an example of two distributions p and q for which $D(p||q)$ is not equal to $D(q||p)$. Give the two distributions and show the explicit computation of the KL divergences. Simple toy distributions (e.g. over heads, tail) are fine.

Problem 2.2 (5 points)
Suppose we are given a data set $\mathbf{x} = (x_1, \ldots, x_N)$, representing N *iid* (independent and identically distributed) observations of the scalar variable x, which follow a Gaussian distribution:

$$p(x) \sim \mathcal{N}(\mu, \sigma^2)$$

a) (2 points)
Find the *maximum likelihood* estimates μ_{ML} and σ_{ML} of the distribution parameters.

b) (3 points)
Show that μ_{ML} is *unbiased* and σ_{ML} is *biased*. Next, replace μ_{ML} with the true value μ in the maximum likelihood estimator σ_{ML} and verify that the resulting estimator $\hat{\sigma}_{ML}$ is *unbiased*.

Bonus Question)
Find the *maximum a posteriori* estimate of the mean μ_{MAP}, given the prior distribution:

$$p(\mu) \sim \mathcal{N}(0, \sigma_m^2)$$

Total points: 10