Machine Learning	Prof. Matthias Hein
Exercise Sheet 2 - 6.5.2021 (12 points)	due: 13.5.2021

Exercise 4 - Maximum Likelihood and Maximum A Posteriori Estimation

- **(3 Points)** Consider the regression problem where the input $X \in \mathbb{R}^d$ and the output $Y \in \mathbb{R}$. Assume that the likelihood is specified in terms of the unknown parameter $w \in \mathbb{R}^d$ as

$$p(y|x,w) = \frac{1}{\sqrt{2\pi g(x)}} e^{-\frac{(y-\langle w,x\rangle)^2}{2g(x)}},$$

where $g : \mathbb{R}^d \to \mathbb{R}_+^*$ is a known positive function. Compute the maximum likelihood estimator of w.

- **(3 points)** Further we have a prior distribution on w:

$$p(w) = \frac{1}{(2\pi)^{\frac{d}{2}} (\prod_{i=1}^d \lambda_i)^{\frac{1}{2}}} e^{-\langle w,\Lambda^{-1}w\rangle}.$$

where Λ is a diagonal matrix with the diagonal entries given by $\frac{1}{\lambda_i}$ ($\lambda_i > 0$ for all $i \in \{1, ..., d\}$). We are given an i.i.d. training sample $(x_i, y_i)_{i=1}^n$, which we assume to be additionally conditionally independent given the model. We impose the condition that w is independent of x. Use this first to show that

$$p(y, x \,|\, w) = p(y \,|\, w, x)p(x).$$

What is the maximum a posteriori (MAP) estimator of w ?

Exercise 5 - ML and MAP estimators

Consider the two r.v., representing two sensors estimating the same value θ,

$$\begin{aligned} A =& \theta + \epsilon_1, \qquad \epsilon_1 \sim \mathcal{N}(0, \sigma_1^2) \\ B =& \theta + \epsilon_2, \qquad \epsilon_2 \sim \mathcal{N}(0, \sigma_2^2), \end{aligned}$$

with ϵ_1 and ϵ_2 independent, and their realizations $(a_1, \dots, a_n)$ and $(b_1, \dots, b_n)$ (in practice $n = 1$ and we need an estimation of θ).

- **(3 points)** Compute the MLE of θ, using the information from both sensors and assuming σ_1, σ_2 known.
- **(3 points)** If additionally we assume a prior

$$p(\theta) = \mathcal{N}(\mu_P, \sigma_P^2),$$

which is the MAP estimator of θ? With which σ_P would we get $\hat{\theta}_{ML} = \hat{\theta}_{MAP}$?