due on Monday, 18 May 2020, 10am sharp

Probabilistic Machine Learning

Exercise Sheet #4

Corona Virus

1. **EXAMple: Gaussian inference**

   Consider the Gaussian random variable $\boldsymbol{w} \in \mathbb{R}^F$ with probability density function

   $$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}; \boldsymbol{\mu}, \Sigma) \qquad \text{with} \qquad \boldsymbol{\mu} \in \mathbb{R}^F \text{ and symmetric positive definite } \Sigma \in \mathbb{R}^{F \times F}.$$

   You have access to data $\boldsymbol{y} \in \mathbb{R}^N$ assumed to be generated from $\boldsymbol{w}$ through a linear map $\Phi \in \mathbb{R}^{F \times N}$ according to the likelihood

   $$p(\boldsymbol{y} \mid \boldsymbol{w}) = \mathcal{N}(\boldsymbol{y}; \Phi^\intercal \boldsymbol{w}, \Lambda),$$

   where $\Lambda \in \mathbb{R}^{N \times N}$ symmetric positive definite. What is

   (a) the pdf of the *marginal* $p(\boldsymbol{y}) = \int p(\boldsymbol{y} \mid \boldsymbol{w}) p(\boldsymbol{w})\, d\boldsymbol{w}$

   (b) the pdf of the *posterior* $p(\boldsymbol{w} \mid \boldsymbol{y})$?

2. **Theory Question: Least Squares**

   Consider the model defined in Exercise 1 (see above), for the special case $\Lambda = \sigma^2 I$ with $\sigma^2 \in \mathbb{R}_+$ (that is, iid. observation noise).

   (a) Show that the **maximum likelihood estimator** for $\boldsymbol{w}$ is given by the **ordinary least-squares** estimate

   $$\boldsymbol{w}_{\text{ML}} = (\Phi\Phi^\intercal)^{-1}\Phi\boldsymbol{y}$$

   To do so, use the explicit form of the Gaussian pdf to write out $\log p(\boldsymbol{y} \mid \boldsymbol{w})$, take the gradient with respect to the elements $[\boldsymbol{w}]_i$ of the vector $\boldsymbol{w}$ and set it to zero. If you find it difficult to do this in vector notation, it may be helpful to write out $[\Phi^\intercal \boldsymbol{w}]_j = \sum_i [\boldsymbol{w}]_i [\Phi]_{ij}$ Calculate the derivative of $\log p(\boldsymbol{y} \mid \boldsymbol{w})$ with respect to $[\boldsymbol{w}]_i$, which is scalar. Setting that to zero, you can bring it to a form $\boldsymbol{v}^\intercal [\Phi]_{i:} = 0$ (where $[\Phi]_{i:}$ is the i-th row of Φ) for some vector $\boldsymbol{v}(\boldsymbol{w})$ that is identical for all i, and thus, stacking up the columns of Φ again, we have $\boldsymbol{v}^\intercal \Phi = 0$. Solving that equation for $\boldsymbol{w}$ yields the desired result.

   (b) By an analogous computation on the posterior $p(\boldsymbol{w} \mid \boldsymbol{y})$, show that the **maximum a-posteriori estimator** is identical to the posterior mean $\boldsymbol{w}_{\text{MAP}} = \mathbb{E}_{p(\boldsymbol{w}|\boldsymbol{y})}(\boldsymbol{w})$ from above. This result shows that, for the particular choice $\boldsymbol{\mu} = 0, \Sigma = I_F$, $\Lambda = \sigma^2 I$, the posterior mean is the **ℓ_2-regularized least-squares** estimator.

3. **Practical Question**

   In the past weeks the world experienced a pandemic unprecedented in our lifetime. Uncertainty and limited data characterized the early phases of the outbreak, precisely the time in which in hindsight decisive action was most crucial. In this exercise you get to try and answer some basic questions about the virus yourself. You will build a simple probabilistic model to estimate the mortality rate of COVID-19 based on temporal data. Inference in this model will be performed using Monte-Carlo methods. On ILIAS you can find a `jupyter` notebook that describes the exercise in more detail and how to obtain the data.

   Note: Although the exercise uses real data, it does not aspire to satisfy the standards of epidemiology or public policy making. It is deliberately simple and designed to be feasible within the scope of this lecture course. Do not mistake it for a scientific analysis.