Machine Learning Exercise Sheet 3

Probabilistic Inference

Exercise sheets consist of two parts: homework and in-class exercises. You solve the homework exercises on your own or with your registered group and upload it to Moodle for a possible grade bonus. The in-class exercises will be solved and discussed during the tutorial along with some difficult and/or important homework exercises. You do not have to upload any solutions of the in-class exercises.

In-class Exercises

Consider the probabilistic model

$$\begin{aligned} \mathrm{p}(\mu \mid \alpha) &= \mathcal{N}(\mu \mid 0, \alpha^{-1}) = \sqrt{\frac{\alpha}{2\pi}} \exp\left(-\frac{\alpha}{2}\mu^2\right) \\ \mathrm{p}(x \mid \mu) &= \mathcal{N}(x \mid \mu, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right) \end{aligned}$$

and a set of observations $\mathcal{D} = \{x_1, ..., x_N\}$ consisting of N samples $x_i \in \mathbb{R}$.

Note: We parametrize $\mu \mid \alpha$ with the *precision* parameter $\alpha = 1/\sigma^2$ instead of the usual variance σ^2 because it leads to a nicer solution.

Problem 1: Derive the maximum likelihood estimate μ_{MLE}. Show your work.

Problem 2: Derive the maximum a posteriori estimate μ_{MAP}. Show your work.

Problem 3: Does there exist a prior distribution over μ such that $\mu_{\text{MLE}} = \mu_{\text{MAP}}$? Justify your answer.

Problem 4: Derive the posterior distribution $\mathrm{p}(\mu \mid \mathcal{D}, \alpha)$. Show your work.

Problem 5: Derive the posterior predictive distribution $\mathrm{p}(x_{new} \mid \mathcal{D}, \alpha)$. Show your work.

Homework

Optimizing Likelihoods: Monotonic Transforms

Usually we maximize the *log-likelihood*, $\log \mathrm{p}(x_1, \ldots, x_n \mid \theta)$ instead of the likelihood. The next two problems provide a justification for this.

In the lecture, we encountered the likelihood maximization problem

$$\arg\max_{\theta \in [0,1]} \theta^t (1-\theta)^h,$$

where t and h denoted the number of tails and heads in a sequence of coin tosses, respectively.

Problem 6: Compute the first and second derivative of this likelihood w.r.t. θ. Then compute first and second derivative of the log-likelihood $\log \theta^t (1-\theta)^h$.

Problem 7: Show that for *any* differentiable, positive function $f(\theta)$ every local maximum of $\log f(\theta)$ is also a local maximum of $f(\theta)$. Considering this and the previous exercise, what is your conclusion?

Properties of MLE and MAP

Problem 8: Consider a Bernoulli random variable X and suppose we have observed m occurrences of $X = 1$ and l occurrences of $X = 0$ in a sequence of $N = m + l$ Bernoulli experiments. We are only interested in the number of occurrences of $X = 1$—we will model this with a Binomial distribution with parameter θ. A prior distribution for θ is given by the Beta distribution with parameters a, b. Show that the posterior *mean* value $\mathbb{E}[\theta \mid \mathcal{D}]$ (not the MAP estimate) of θ lies between the prior mean of θ and the maximum likelihood estimate for θ.

To do this, show that the posterior mean can be written as λ times the prior mean plus $(1-\lambda)$ times the maximum likelihood estimate, with $0 \leq \lambda \leq 1$. This illustrates the concept of the posterior mean being a compromise between the prior distribution and the maximum likelihood solution.

The probability mass function of the Binomial distribution for some $m \in \{0, 1, \ldots, N\}$ is

$$\mathrm{p}(x = m \mid N, \theta) = \binom{N}{m} \theta^m (1-\theta)^{N-m}.$$

Hint: Identify the posterior distribution. You may then look up the mean rather than computing it.

Problem 9: Consider the following probabilistic model

$$\mathrm{p}(\lambda \mid a, b) = \mathrm{Gamma}(\lambda \mid a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp(-b\lambda)$$
$$\mathrm{p}(x \mid \lambda) = \mathrm{Poisson}(x \mid \lambda) = \frac{\lambda^x \exp(-\lambda)}{x!}$$

where $a \in (1, \infty)$ and $b \in (0, \infty)$. We have observed a single data point $x \in \mathbb{N}$. Derive the maximum a posteriori (MAP) estimate of the parameter λ for the above probabilistic model. Show your work.

Upload a single PDF file with your homework solution to Moodle by 10.11.2021, 23:59 CET. We recommend to typeset your solution (using LaTeX or Word), but handwritten solutions are also accepted. If your handwritten solution is illegible, it won't be graded and you waive your right to dispute that.

Programming Task

Problem 10: Download the notebook `exercise_03_notebook.ipynb` from Moodle. Fill in the missing code and follow the instructions in the notebook to append the solution to your PDF submission.

Note: We suggest that you use Anaconda for installing Python and Jupyter, as well as for managing packages. We recommend that you use Python 3.

For more information on Jupyter notebooks, consult the Jupyter documentation. Instructions for converting the Jupyter notebooks to PDF are provided within the notebook.