Machine Learning
Prof. S. Harmeling

Wednesday, 27. October 2021
DUE 23:55 Tuesday, 02. November 2021

Exercise set #3

Please submit your solutions in teams of two using the sciebo file-drop folder. The link is available in ILIAS. For the formatting please stick to the `submission_guideline.pdf` that you can find on sciebo. In the case of multiple uploads we will consider the latest. Uploads after the deadline will be deleted without further notice.

1. **MLEs**
   Derive the maximum likelihood estimators for the following distributions. For this write down the log-likelihood function given n observations $x_1, \ldots, x_n$ and determine the maximum with respect to the parameter.
   (a) Gaussian normal distribution $\mathcal{N}(y|\mu, \sigma^2)$ for μ.
   (b) Exponential distribution with probability density function $f(x|\lambda) = \lambda e^{-\lambda x}$ for $\lambda > 0$ given observations $x_i \geq 0$.
   (c) Gamma distribution with probability density function $g(x|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x^{(\alpha-1)} e^{-\lambda x}$ for parameter λ given observations $x_1, \ldots, x_n \geq 0$ and a known $\alpha > 0$.

   Hint: The derivative of $\log(y)$ is $\frac{1}{y}$.

   45 points

2. **Dirichlet-multinomial model**
   Throwing a (not necessarily fair) K-sided die n times allows us to infer posteriors for the unknown probabilities. The data is $\mathcal{D} = (x_1, \ldots, x_K)$ with x_j being the number of times you have seen side j. Assume a Dirichlet prior (with (hyper-)parameter vector α) for the parameter vector $\theta = (\theta_1, \ldots, \theta_K)$ with $0 \leq \theta_j \leq 1$ and $\sum_j \theta_j = 1$ and a multinomial likelihood for your data, i.e.,

   $$p(\theta) = \mathrm{Dir}(\theta|\alpha) \qquad\qquad p(\mathcal{D}|\theta) = \mathrm{Mu}(x|n, \theta)$$

   Show that the posterior is also Dirichlet, i.e., show

   $$p(\theta|\mathcal{D}) = \mathrm{Dir}(\theta|\alpha + x)$$

   Hint: You do not have to calculate the normalization constant, i.e., prove that the posterior is proportional to a Dirichlet distribution with parameter $\alpha + x$.

   30 points

3. **Inference for a difference in proportions (programming task)**
   Recall the story from the lecture "Two sellers at Amazon have the same price. One has 90 positive and 10 negative reviews. The other one 2 positive and 0 negative. Who should you buy from?" Write down the posterior probabilities about the reliability (as in the lecture).
   (a) Calculate $p(\theta_1 > \theta_2|\mathcal{D}_1, \mathcal{D}_2)$ using quadrature, e.g., by using the function `dblquad` from `scipy.integrate`.
   (b) Calculate $p(\theta_1 > \theta_2|\mathcal{D}_1, \mathcal{D}_2)$ using Monte Carlo integration[1]. You can generate Beta distributed samples with the function `scipy.stats.beta.rvs(a,b,size)`.

   25 points

[1] `https://en.wikipedia.org/wiki/Monte_Carlo_integration`