Machine Learning 2 **Due: 23:59, September 9, 2019**

Homework 1

Instructors: Adeel Pervez, Pim de Haan, Noud de Kroon, David Zhang
Email: a.a.pervez@uva.nl, pimdehaan@gmail.com, a.a.w.m.dekroon@uva.nl, w.d.zhang@uva.nl

You are allowed to discuss with your colleagues but you should write the answers in your own words. If you discuss with others, write down the name of your collaborators on top of the first page. No points will be deducted for collaborations. If we find similarities in solutions beyond the listed collaborations we will consider it as cheating. We will not accept any late submissions. The solutions to the previous homework set will be put on canvas by the end of the day of the hand-in date. $\star$ denotes bonus exercise. You earn 1 point for solving each bonus exercise. All bonus points earned will be added to your total homework points.

Problem 1. (1 pt) Consider two random vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^n$ having Gaussian distribution $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_\mathbf{x})$ and $p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_\mathbf{z}, \boldsymbol{\Sigma}_\mathbf{z})$. Consider random vector $\mathbf{y} = \mathbf{x} + \mathbf{z}$. Derive mean and covariance of $p(\mathbf{y})$. What is the covariance of $\mathbf{y}$ if you assume that $\mathbf{x}$ and $\mathbf{z}$ are independent?

Problem 2. (0.5+0.5+1.5+0.5 = 3 pts) Consider a D-dimensional Gaussian random variable $\mathbf{x}$ with distribution $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in which the covariance $\boldsymbol{\Sigma}$ is known. Given a set of N i.i.d. observations $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Assume that $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. [Hint: you may directly use results from Bishop]

1. Write down the likelihood of the data $p(\mathcal{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$;
2. Write down the form of the posterior $p(\boldsymbol{\mu}|\mathcal{X}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ (you do not need to normalize the probability distribution by calculating the evidence).
3. Show that $p(\boldsymbol{\mu}|\mathcal{X}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ is a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$ and find the values of $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ (hint: use "completing the square")
4. Derive the maximum a posteriori solution for $\boldsymbol{\mu}$;

Problem 3. (0.5 + 0.5 + 0.5 = 1.5 pts) Tossing a biased coin with probability that it comes up heads is μ. [Hint: you may use results from Bishop]

1. We toss the coin 3 times and it all comes up with heads. How likely is that in the next toss, the coin comes up with head according to MLE?
2. Suppose that the prior $\mu \sim \text{Beta}(\mu|a, b)$. What is the probability that the coin comes up with head in the 4th toss?
3. Suppose that we observe m times that the coin lands heads and l times that it lands tails. Show that the posterior mean $\mathrm{E}[\mu|\mathcal{D}]$ (see Bishop 2.19) lies between the prior mean and μ_{MLE}.

Problem 4. (2 + 1 + 0.5 = 3.5 pts) Consider the following distributions:

(i) $\text{Pois}(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$

(ii) $\mathrm{Gam}(\tau|a,b) = \frac{1}{\Gamma(a)} b^a \tau^{a-1} e^{-b\tau}$

(iii) $\mathrm{Cauchy}(x|\gamma,\mu) = \frac{1}{\pi\gamma} \frac{1}{1+(\frac{x-\mu}{\gamma})^2}$

(iv) $\mathrm{vonMises}(x|\kappa,\mu) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\mu)}$

Answer the following questions:

1. Are the above distributions members of an exponential family. If yes, then (a) cast them in exponential form (Bishop eq. 2.194) with a minimum numbers of parameters, (b) derive their sufficient statistics.

2. Derive the first moment about zero (i.e. the mean) and the second moment about the mean (i.e. the variance) of the distributions (i) and (ii).

3. Does the Poisson distribution have a conjugate prior? Derive the conjugate prior, if the answer is "yes".

Problem 5⋆. (1 pt) Derive mean, covariance, and mode of multivariate Student's t-distribution.