Machine Learning 2	**Due: April 11, 2017**

Homework 1

Instructor: Stephan Bongers, Tineke Blom   Email: s.r.bongers@uva.nl, t.blom2@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 under any circumstances. The solutions to the previous homework will be handed out in the class at the beginning of the next homework session. After this point, late submissions will be automatically graded zero.
$\star$ 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. 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 $p(\mathbf{y})$.

Problem 2. 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 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 use results from Bishop]

1. Write down the likelihood of the data $p(\mathcal{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$;
2. Write down the posterior $p(\boldsymbol{\mu}|\mathcal{X}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$;
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 $\boldsymbol{\Sigma}_N$;
4. Derive the maximum a posterior solution for $\boldsymbol{\mu}$;

Consider now the univariate ($D = 1$) case, where Σ is equal to σ^2.

5. Derive expressions for sequential update of μ_N and σ_N^2 from the results you found in 3;
6. Derive the same results (as in 5) starting from the posterior distribution $p(\mu|x_1, \ldots, x_{N-1})$, and multiplying by the likelihood function $p(x_N|\mu) = \mathcal{N}(x_N|\mu, \sigma^2)$.

Problem 3.

1. Show that the product of two Gaussians gives another (un-normalized) Gaussian
$$\mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{A})\mathcal{N}(\mathbf{x}|\mathbf{b}, \mathbf{B}) = K^{-1}\mathcal{N}(\mathbf{x}|\mathbf{c}, \mathbf{C})$$
where $\mathbf{c} = \mathbf{C}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b})$ and $\mathbf{C} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}$.
2. Using the *matrix inversion lemma*, also known as the the Woodbury, Sherman & Morrison formula:
$$(\mathbf{Z} + \mathbf{U}\mathbf{W}\mathbf{V}^\intercal)^{-1} = \mathbf{Z}^{-1} - \mathbf{Z}^{-1}\mathbf{U}(\mathbf{W}^{-1} + \mathbf{V}^\intercal\mathbf{Z}^{-1}\mathbf{U})^{-1}\mathbf{V}^\intercal\mathbf{Z}^{-1} \tag{1}$$
Proof that $\mathbf{C} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{A} - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} = \mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$

3. Show that

$$K^{-1} = (2\pi)^{-D/2}|\mathbf{A}+\mathbf{B}|^{-1/2}\exp\big(-\frac{1}{2}(\mathbf{a}-\mathbf{b})^{\mathsf{T}}(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{a}-\mathbf{b})\big) \tag{2}$$

Problem 4. Tossing a biased coin with probability that it comes up heads is μ. [Hint: use 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 lies between the prior mean and μ_{MLE}.

Problem 5. Verify that the Poisson and Gamma distributions can be cast in exponential family form (Bishop eq. 2.194) and derive expressions for $\boldsymbol{\mu}, \mathbf{u}(\mathbf{x}), h(\mathbf{x})$ and $g(\boldsymbol{\mu})$. Derive the mean from $g(\boldsymbol{\mu})$ for both distributions.

Problem 6$\star$. Derive mean, covariance, and mode of multivariate Student's t-distribution.