# TDA 231 Machine Learning: Homework 1

Goal: Maximum likelihood estimation (MLE), Maximum a posteriori (MAP)
Grader: Jonatan Kilhamn

Due Date: January 30, 2017

**General guidelines:**

1. All solutions to theoretical problems, and discussion regarding practical problems, should be submitted in a single file named *report.pdf*
2. All matlab files have to be submitted as a single zip file named *code.zip*.
3. The report should clearly indicate your name, personal number and email address
4. All datasets can be downloaded from the course website.
5. All plots, tables and additional information should be included in *report.pdf*

## 1 Theoretical problems

**Problem 1.1** [Maximum likelihood estimator (MLE), 4 points]

Consider a dataset $\mathbf{x}_1, \ldots, \mathbf{x}_n$ consisting of i.i.d. observations generated from a *spherical* multivariate Gaussian distribution $N(\boldsymbol{\mu}, \sigma^2 I)$, where $\boldsymbol{\mu} \in \mathbb{R}^p$, $I$ is the $p \times p$ identity matrix, and $\sigma^2$ is a scalar. Derive the maximum likelihood estimator for $\sigma$.

**Problem 1.2** [Posterior distributions, 6 points]

Consider dataset $\mathbf{x}_1, \ldots, \mathbf{x}_n$ consisting of i.i.d. observations generated from a *spherical* multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \sigma^2 I)$, where $\boldsymbol{\mu} = [\mu_1, \mu_2]^\top \in \mathbb{R}^2$, $I$ is the $2 \times 2$ identity matrix, and $\sigma^2$ is a scalar. The probability distribution of a point $\mathbf{x} = [x_1, x_2]^\top$ is given by

$$P(\mathbf{X} = \mathbf{x} | \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left( -\frac{(\mathbf{x} - \boldsymbol{\mu})^\top (\mathbf{x} - \boldsymbol{\mu})}{2\sigma^2} \right) .$$

We assume that $\sigma^2$ has an *inverse-gamma* prior distribution given by

$$P(\sigma^2 = s | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{-\alpha - 1} \exp\left( -\frac{\beta}{s} \right) , \tag{1}$$

where $\alpha$ and $\beta$ are parameters and $\Gamma(\cdot)$ is the gamma function given by $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

(a) Derive the posterior distribution $p(\sigma^2 = s|\mathbf{x}_1, \ldots, \mathbf{x}_n; \alpha, \beta)$. (HINT: inverse-gamma distribution is conjugate prior to spherical Gaussian distribution when mean is known).

(b) Choose $\boldsymbol{\mu}$ to be the empirical mean of the data, $\bar{\mathbf{x}}$ and consider two separate models of the same form as in (a), but having different parameters for the inverse-gamma prior

- $\alpha_A$ and $\beta_A$ (Model $M_A$)
- $\alpha_B$ and $\beta_B$ (Model $M_B$)

Write the expression for the Bayes factor for these models, without computing the integrals.

(c) Make the assumptions:

- $P(M_A) = P(M_B) = \frac{1}{2}$, and
- Use the MAP estimate for $\sigma^2$.

Derive the expression for the Bayes Factor under the above assumptions.

## 2 Practical problems

**Useful matlab functions:**

- *General:* arrayfun, cellfun, crossvalind, reshape, (anonymous functions using @), min, mat2cell, cell2mat
- *Plotting:* plot, scatter, legend, hold, imshow, subplot, grid, title, saveas

**Problem 2.1** [Spherical Gaussian estimation, 5 points]

Consider a dataset consisting of i.i.d. observations generated from a spherical Gaussian distribution $N(\boldsymbol{\mu}, \sigma^2 I)$, where $\boldsymbol{\mu} \in \mathbb{R}^p$, $I$ is the $p \times p$ identity matrix, and $\sigma^2$ is a scalar.

(a) Write the mathematical expression for the MLE estimators for $\boldsymbol{\mu}$ and $\sigma$ in above setup.

(b) Implement a matlab function *sge()* that estimates the mean $\boldsymbol{\mu}$ and variance $\sigma^2$ from the given data, using the skeleton code provided below (or *sge.m* on the website).

```
function [mu, sigma] = sge(x)
%
% SGE Mean and variance estimator for spherical Gaussian distribution
%
% x      : Data matrix of size n x p where each row represents a
%          p-dimensional data point e.g.
%              x = [2 1;
%                   3 7;
%                   4 5 ] is a dataset having 3 samples each
%                   having two co-ordinates.
%
% mu     : Estimated mean of the dataset [mu_1 mu_2 ... mu_p]
% sigma  : Estimated standard deviation of the dataset (number)
%

YOUR CODE GOES HERE
```

(c) Implement a function which takes as input a two-dimensional dataset $x$ (as described above); and draws, on the same plot, the following:

1. A scatter plot of the original data $x$,
2. Circles with center $\boldsymbol{\mu}$ and radius $r = k\sigma$ for $k = 1, 2, 3$ where $\boldsymbol{\mu}$ and $\sigma^2$ denote the mean and variance estimated using *sge()*.
3. Legend for each circle indicating the fraction of points (in the original dataset) that lie outside the circle boundary.

(d) Run your code on the dataset dataset1.mat. Submit the resulting plot as well as your implementation.

**Problem 2.2** [MAP estimation, 5 points]

Consider a dataset consisting of i.i.d. observations generated from a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \sigma^2 I)$, where $\boldsymbol{\mu} = [\mu_1, \mu_2]^\top \in \mathbb{R}^2$, $I$ is the $2 \times 2$ identity matrix, and $\sigma^2$ is a scalar. We will now explore the Bayesian approach to estimation of $\sigma^2$ *under the assumption that the mean* $\boldsymbol{\mu}$ *is known.* The probability distribution of a point $\mathbf{x} = [x_1, x_2]^\top$ is given by

$$P(\mathbf{X} = \mathbf{x}|\sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^\top(\mathbf{x}-\boldsymbol{\mu})}{2\sigma^2}\right)$$

We assume $\sigma^2$ has *inverse-gamma* prior distribution given by

$$P(\sigma^2 = s|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{-\alpha-1} \exp\left(-\frac{\beta}{s}\right) \tag{2}$$

where $\alpha$ and $\beta$ are parameters and $\Gamma(\cdot)$ is the gamma function given by $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$.

For the following tasks, use dataset1.mat, provided on the web page.

(a) On the same plot, show the prior and posterior distributions for $\sigma$ with parameters $\alpha = 1$ and $\beta = 1$. Generate a second plot with $\alpha = 10$ and $\beta = 1$. What do you observe? (HINT: You might want to check out the "`log-sum-exp` trick")

(b) Choose $\boldsymbol{\mu}$ to be the empirical mean and consider two separate models (having different parameters)

- $\alpha_a = 1$ and $\beta_a = 1$ (Model $M_A$)
- $\alpha_b = 10$ and $\beta_b = 1$ (Model $M_B$)

Compute analytically the expression for the MAP estimate for both models in terms of posterior parameters $\alpha_1, \beta_1$. Report the numerical values of the MAP estimates for the two models.

(c) Now we ask "Which is the better model?". Write the expression for the Bayes factor under two assumptions:

- $P(M_A) = P(M_B) = \frac{1}{2}$, and
- Use the MAP estimate for $\sigma^2$.

Compute and report the Bayes factor for the two models using the MAP estimate above and, consequently, state which is the better model.