EE-613: Probabilistic Generative Models, Lab 2, Part 1

Fall 2019

1 Maximum-A-Posteriori adaptation of a multivariate Gaussian

The goal is to study the estimation of the parameters ($\boldsymbol{\mu}$,Σ) using the Maximum A Posteriori (MAP) principle. Comparison with the Maximum-Likelihood estimator and the effect of the prior parameters will be studied.

Let's assume that we are give a set of i.i.d. observations $\mathcal{X} = \{\mathbf{x}_i, i = 1 \dots N\}$ that follow a multivariate Gaussian law $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \Sigma)$, where the data points are of dimension d. It has been shown that the conjugate prior for the parameter $(\boldsymbol{\mu}, \Sigma)$ of this multivariate Gaussian likelihood distribution is the Normal-Inverse Wishart distribution:[1]

$$p(\boldsymbol{\mu}, \Sigma|\mathbf{m}, \tau, V, \nu) = \mathcal{NIW}(\boldsymbol{\mu}, \Sigma|\boldsymbol{\beta}) = \mathcal{NIW}(\boldsymbol{\mu}, \Sigma|\mathbf{m}, \tau, V, \nu) = p(\boldsymbol{\mu}|\Sigma, \mathbf{m}, \tau)p(\Sigma|V, \nu) \quad (1)$$

where $\boldsymbol{\beta} = (\mathbf{m}, \tau, V, \nu)$ denotes the set of parameters defining the prior distribution, and the prior on the mean given the covariance is given by a normal distribution:

$$p(\boldsymbol{\mu}|\Sigma, \mathbf{m}, \tau) = \mathcal{N}(\boldsymbol{\mu}|\mathbf{m}, \frac{1}{\tau}\Sigma), \quad (2)$$

and the conjugate prior of the covariance is an inverse Wishart distribution given by:

$$p(\Sigma|V, \nu) = \mathcal{IW}(\Sigma|V, \nu) = B|\Sigma|^{-(\nu+d+1)/2} \exp\left(-\frac{1}{2}tr(V\Sigma^{-1})\right) \quad (3)$$

where B is some normalization constant. According to the conjugacy definition, the posterior for the Gaussian parameters also follows a Normal-Inverse Wishart distribution given by:

$$p(\boldsymbol{\mu}, \Sigma|\mathcal{X}, \boldsymbol{\beta}) = \mathcal{NIW}(\boldsymbol{\mu}, \Sigma|\mathbf{m}_{new}, \tau_{new}, V_{new}, \nu_{new}) \quad (4)$$

with

$$\mathbf{m}_{new} = \frac{\tau\mathbf{m} + N\bar{\mathbf{x}}}{\tau + N}, \tau_{new} = \tau + N, \nu_{new} = \nu + N, V_{new} = V + NS + \frac{\tau N}{\tau + N}(\mathbf{m} - \bar{\mathbf{x}})(\mathbf{m} - \bar{\mathbf{x}})' \quad (5)$$

where $\bar{\mathbf{x}}$ denotes the sample mean $\bar{\mathbf{x}} = \frac{1}{N}\sum_i \mathbf{x}_i$, and $S = \frac{1}{N}\sum_i(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. Finally, from this distribution, the MAP estimates for the mean and covariance are given by

$$\boldsymbol{\mu}_{MAP} = \frac{\tau\mathbf{m} + N\bar{\mathbf{x}}}{\tau + N} \quad (6)$$

$$\Sigma_{MAP} = \frac{V + NS + \frac{\tau N}{\tau+N}(\mathbf{m} - \bar{\mathbf{x}})(\mathbf{m} - \bar{\mathbf{x}})'}{\nu - d + N} \quad (7)$$

[1] The wikipedia website provides such information regarding all types of distributions, Wikipedia.

Questions

You are given a jupyter notebook part-1.ipynb containing the code to use and to complete. Follow the steps of this notebook while at the same time reading and answering the questions below.

(a) Draw the graphical model of the $(\boldsymbol{\mu}, \Sigma, \mathcal{X})$ variables (you can use the plate notation). Remind the conjugacy principle. Remind the MAP principle and thus to what correspond the above MAP estimates w.r.t. distributions provided above.

(b) Programming. The notebook contains missing parts. The data generation function is written and some parameter settings for the wisharts are proposed within a function. In addition, there is a function `show_results` which can be used to illustrate different gaussian parameters (it shows filled ellipses centered around the mean and at one standard deviation to visualize the covariance). You are asked to program:

- a function which performs Maximum Likelihood estimation:

```
def ml_gauss_param(X):
    # Perform the Maximum Likelihood adaptation of Gaussian parameter.
    # Input: X: data points.
    # Output: mu,sigma: estimated Gaussian parameters.
```

- a function which compute the MAP estimates of the multivariate Gaussian parameters under normal-wishart prior:

```
# MAP adaptation.
def map_gauss_param(X, wparam):
    # Perform the MAP adaptation of Gaussian parameter using Wishart prior.
    # Input: X: data points.
    #        wparam: structure containing the Wishart parameters.
    # Output: mu, sigma: estimated Gaussian parameters.
```

Assuming 100 points are generated, the following default parameters are assumed for the prior: $\mathbf{m} = (-10, -10)'$, $\tau = 100$, $\nu = 100$, and $V = 100 \times 30 I_2$ where I_2 denotes the 2×2 identity matrix.

(c) By inspecting the prior distribution on the mean and its MAP estimate, provide an interpretation of the $\mathbf{m}$ and τ parameters. Illustrate your answer by testing different values of τ, e.g $5, 100, 500$. What can you say about the special value $\tau = N$?

(d) Comment on the consequences of varying τ (e.g. towards 0, or towards values $\gg N$) on the MAP estimate of the covariance Σ_{MAP}. Why does it make sense?

(e) The MAP estimation of the covariance is also controlled through the Wishart parameters V and ν. It can be shown that the expected covariance Σ_E under the inverse Wishart law is given by $\frac{1}{\nu-d-1}V$ (write down the equation corresponding to what this sentence mean). To *manually* define the Wishart parameters V and ν, it might be thus more convenient to define first Σ_E (as it is more easily interpretable w.r.t. to our data) as well as ν, and then set V accordingly.

Try different values of ν (e.g. 25,100,1000) (use the above method for setting V, i.e. keep the same expected covariance matrix Σ_E), and visualize the obtained results. Given these results and looking at the formula of the MAP estimate of the covariance, explain the qualitative role of this parameter.