Generative Models Exercises

Dr. Jean Marc Odobez (https://www.idiap.ch/~odobez)
Dr. Michael Villamizar (https://www.idiap.ch/~mvillamizar/)
Anshul Gupta (https://www.idiap.ch/~agupta/)

Outline

Part 2: ML and MAP Estimation

Outline

Part 2: ML and MAP Estimation

ML and MAP Estimation

Goal

This laboratory notebook focuses on understanding the Maximum A Posteriori (MAP) estimation principle, as compared to the more traditional Maximum Likelihood (ML) case. This principle is illustrated here using the multivariate Gaussian distribution as the (data) likelihood distribution.

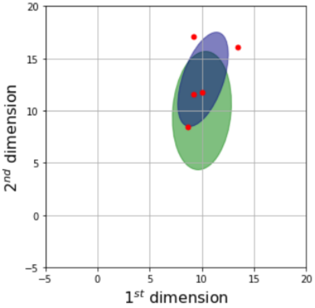
More precisely, the goal is to study the estimation of the parameters of the mean ($\boldsymbol{\mu}$) of such a distribution using this MAP principle, as well as understand the effect of the parameters of the prior distribution on the result.

ML and MAP Estimation

Dataset

The data is sampled from a Gaussian:

- N samples
- 2 dimensions



ML and MAP Estimation

Exercise 1: ML Estimation

Complete the code in the following function to return the Maximum Likelihood estimation of the parameters of a Gaussian.

Solution: Code

```
mean = np.mean(X, axis=0)
cov = np.cov(X.T, ddof=0)
```

ML and MAP Estimation

Exercise 2: MAP Estimation

Complete the code in the following function to return the parameters of the posterior distribution over the mean of the data distribution.

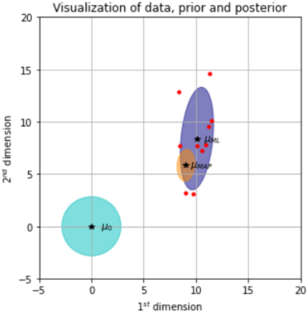
Solution: Code

```
priormean = priorparam['mean']
priorcov = priorparam['cov']
N, d = np.shape(X)

x_bar = np.mean(X,axis=0)
P_prior = np.linalg.inv(priorcov)
P_data = np.linalg.inv(CovData)

cov = np.linalg.inv(P_prior + N*P_data)
mean = np.matmul(cov, (np.matmul(P_prior, priormean) \
                  + N*np.matmul(P_data, x_bar)))
```

ML and MAP Estimation

Visualization

- the data (red points)
- the mean and covariance of the data (purple ellipse)
- the prior distribution on the mean (its mean and covariance; cyan ellipse)
- the posterior distribution on the mean (its mean and covariance; orange ellipse)



ML and MAP Estimation

Exercise 3: Studying the effect of the N value

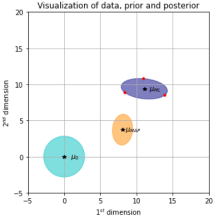
Complete the code below. For each value of N, generate N samples, compute the parameters of the mean posterior distribution, and visualize the different distributions (the prior, the posterior, and the data gaussian) using the 'utils.show_results' function.

Solution: Code

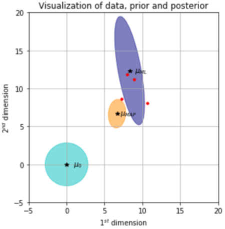
```
for N in N_list:
    X_new, CovData = generate_data(N, d)
    posteriorparam_new = map_gauss_param(X_new, priorparam, CovData)

    utils.show_results(X_new, priorparam, posteriorparam_new)
```

ML and MAP Estimation

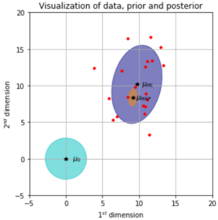
N = 3



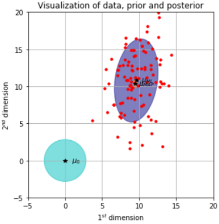
N = 5



N = 20



N = 100



ML and MAP Estimation

Question 1a: Studying the effect of the N value

Observe the estimated means and covariances of the generate data. Remember that all these datasets are sampled from the same Gaussian distribution (run the above cell several times). For which values of N do they fluctuate more/less. Explain.

Solution:

Although all datasets are generated from the same Gaussian, the fluctuate quite a lot, especially for small values of N.

ML and MAP Estimation

Question 1b: Studying the effect of the N value

Visually, how does increasing N affect the values of $\mu_{\mathbf{p}}$ and Σ_p? Explain this with the help of the above equations.

Solution:

As N increases, we see that the mean of the posterior distribution converges to the sample mean. We also see that the covariance of the posterior distribution decreases.

For the covariance, as N increases, the weight of the $N\Sigma^{-1}$ term in Σ_p^{-1} increases. This encourages Σ_p^{-1} to be more similar to $N\Sigma^{-1}$. And as Σ is constant, $N\Sigma^{-1}$ increases and thus Σ_p decreases.

For the mean, $\mu_{\mathbf{p}}$ is a linear combination of $\mu_{\mathbf{0}}$ and $\bar{x}$. So as N increases, the weight of $\bar{x}$ increases. Note however they do not lie on a straight line joining the prior mean to the sample mean.

ML and MAP Estimation

Exercise 3: Studying the effect of the data covariance

Complete the code below. For each value of N and the scaled Σ, generate N samples, compute the parameters of the mean posterior distribution, and visualize the different distributions (the prior, the posterior, and the data gaussian) using the 'utils.show_results' function.

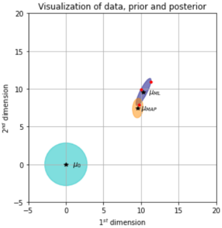
Solution: Code

```
for N in N_list:
    for s in scale:
        CovData_new = CovData * s
        X_new, _ = generate_data(N, d, CovData_new)
        posteriorparam_new = map_gauss_param(X_new, priorparam, CovData_new)

        utils.show_results(X_new, priorparam, posteriorparam_new)
```

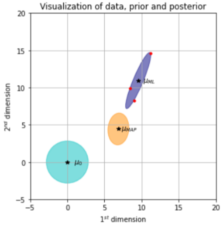
NOTE: generate_data should also use CovData_new

ML and MAP Estimation

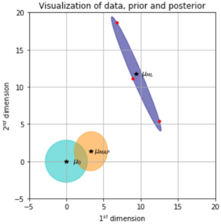

```
N =  3
Data covariance scale factor =  0.2
```



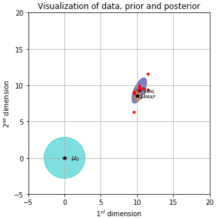

```
N =  3
Data covariance scale factor =  1
```



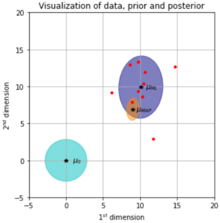

```
N =  3
Data covariance scale factor =  5
```



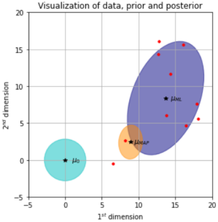

```
N =  10
Data covariance scale factor =  0.2
```




```
N =  10
Data covariance scale factor =  1
```




```
N =  10
Data covariance scale factor =  5
```



ML and MAP Estimation

Question 2: Studying the effect of the data covariance

How does the value of the data covariance affect your observation above? Explain with the help of the posterior equations.

Solution:

We see that the mean of the posterior converges more slowly to $\bar{x}$ for increasing N. We also see that the covariance of the posterior decreases more slowly for increasing N.
For the covariance, as Σ is larger, Σ_p is closer to Σ_0. As N increases, the weight of the $N\Sigma^{-1}$ term increases and this effect is mitigated.
For the mean, as Σ is larger, the weight for $\mu_{\mathbf{0}}$ is higher. As N increases, the weight for $\bar{x}$ increases but the weight is smaller compared to the previous observations due to the larger Σ.

ML and MAP Estimation

Exercise 4: Studying the effect of the data covariance

Complete the code below.
Until now we have assumed knowledge of the covariance Σ used to generate the data but this is often not true in real use-cases. One possiblity is to use the ML estimate Σ_{ML}. Using Σ and Σ_{ML}, for each value of N generate N samples, compute the parameters of the mean posterior distribution, and visualize the different distributions (the prior, the posterior, and the data gaussian) using the 'utils.show_results' function.

ML and MAP Estimation

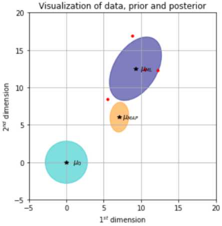
Solution: Code

```
# Using the known data covariance matrix
for N in N_list:
    X_new, CovData = generate_data(N, d)
    posteriorparam = map_gauss_param(X_new, priorparam, CovData)

    utils.show_results(X_new, priorparam, posteriorparam)

# Using the ML estimate of the data covariance matrix
for N in N_list:
    X_new, CovData = generate_data(N, d)
    CovData_ML = ml_gauss_param(X_new)['cov']
    posteriorparam_new = map_gauss_param(X_new, priorparam, CovData_ML)

    utils.show_results(X_new, priorparam, posteriorparam_new)
```

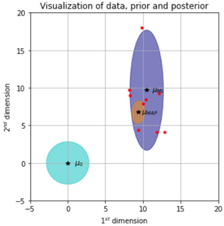
ML and MAP Estimation

Using the known data covariance matrix

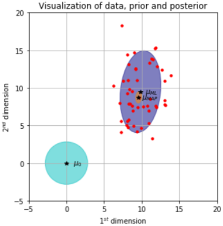
N = 4



N = 10

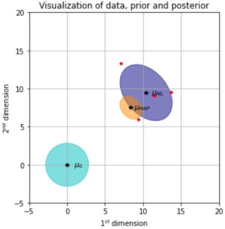


N = 50

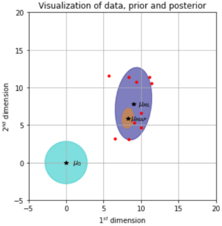


Using the ML estimate of the data covariance matrix

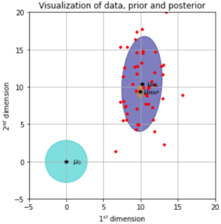
N = 4



N = 10



N = 50



ML and MAP Estimation

Question 3: Studying the effect of the data covariance

What do you observe when using the empirical covariance? Does it reflect the 'Ground truth' about the data (from which random variable they have been generated)? In this view, to which extent can Σ_{ML} serve as a good proxy for Σ? Does the value of N matter, and how?

Solution:

- The posterior uncertainty reflects the estimated covariance (shape, size)
- Although the underlying distribution the data are coming from is still the same
- So, Σ_{ML} can serve as a proxy for Σ, especially as N increases (where it converges to the underlying True covariance).
- Its value is of course very variable, esp. for small N.

ML and MAP Estimation

Exercise 5: Studying the effect of the prior covariance

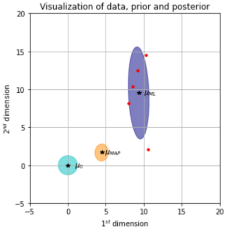
Complete the code below. For each value of N and the scaled Σ_0, generate N samples, compute the parameters of the mean posterior distribution, and visualize the different distributions (the prior, the posterior, and the data gaussian) using the 'utils.show_results' function.

Solution: Code

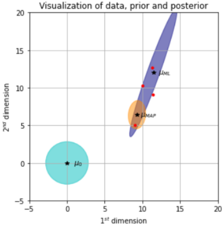
```
priorparam_new['mean'] = priorparam['mean']
for N in N_list:
    for s in scale:
        priorparam_new['cov'] = priorparam['cov'] * s
        X_new, _ = generate_data(N, d)
        posteriorparam_new = map_gauss_param(X_new, priorparam_new, CovData)

        utils.show_results(X_new, priorparam_new, posteriorparam_new)
```

ML and MAP Estimation

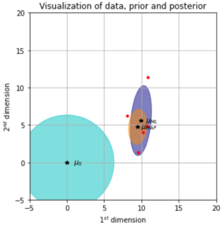
```
N =  5
Prior covariance scale factor =  0.2
```



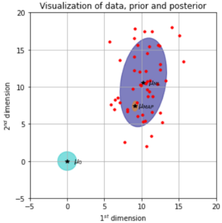

```
N =  5
Prior covariance scale factor =  1
```



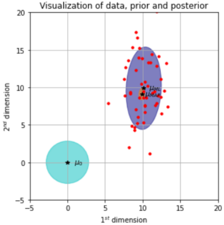

```
N =  5
Prior covariance scale factor =  5
```



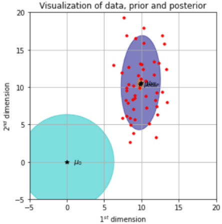

```
N =  50
Prior covariance scale factor =  0.2
```




```
N =  50
Prior covariance scale factor =  1
```




```
N =  50
Prior covariance scale factor =  5
```



ML and MAP Estimation

Question 4: Studying the effect of the prior covariance

Visually, how does the value of the prior covariance affect the posterior parameters $\mu_{\mathbf{p}}$ and Σ_p for different values of N? Explain with the help of the posterior equations.

Solution:

We see that for a given N, $\mu_{\mathbf{p}}$ is closer to $\mu_{\mathbf{0}}$ when Σ_0 is smaller. Also, Σ_p is smaller when Σ_0 is smaller. As N increases, $\mu_{\mathbf{p}}$ moves closer to $\bar{x}$ and Σ_p decreases.
For the covariance, as Σ_0 is larger, Σ_p is closer to $\frac{1}{N}\Sigma$. As N increases, the weight of the $N\Sigma^{-1}$ term further increases and Σ_p decreases.
For the mean, as Σ_0 is larger, the weight for $\bar{x}$ is higher. As N increases, the weight for $\bar{x}$ further increases.

ML and MAP Estimation

Question 5: Studying the effect of the prior covariance

Do you think that the prior is well set? From your observation above, what would you choose for the value of Σ_0?

Solution:

The prior can be set in a better manner.
From the previous study, we can see that choosing a larger Σ_0 results in the $\mu_{\mathbf{p}}$ being closer to $\bar{x}$. Hence, we place more emphasis on the data. Although this results in a larger value for Σ_p, this effect is mitigated with a larger N.

Thank you for your attention!

Dr. Jean Marc Odobez (https://www.idiap.ch/~odobez)
Dr. Michael Villamizar (https://www.idiap.ch/~mvillamizar/)
Anshul Gupta (https://www.idiap.ch/~agupta/)
Idiap Research Institute, Martigny, Switzerland