

CS405 Homework #2

Course: Machine Learning(CS405) - Instructor: Qi Hao

Homework Submission Instructions

Please write up your responses to the following problems clearly and concisely. We require you to write up your responses with A4 paper. You are allowed and encouraged to work together. You may discuss the homework to understand the problem and reach a solution in groups. *However, each student must write down the solution independently.* You must understand the solution well enough in order to reconstruct it by yourself. (This is for your own benefit: you have to take the exams alone.)

- **Written Homeworks.** All calculation problems MUST be written on a pdf document or a word document to bb. Submitting the scan or photo version will **NOT** be accepted.
- **Coding Homeworks.** On online judge. Will be released soon.

Question 1

(a) **[True or False]** If two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is again Gaussian. Similarly, the marginal distribution of either set is also Gaussian

(b) Consider a partitioning of the components of x into three groups x_a , x_b , and x_c , with a corresponding partitioning of the mean vector μ and of the covariance matrix Σ in the form

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_c \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc} \end{pmatrix}.$$

Find an expression for the conditional distribution $p(x_a|x_b)$ in which x_c has been marginalized out.

Question 2

Consider a joint distribution over the variable

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

whose mean and covariance are given by

$$\mathbb{E}[\mathbf{z}] = \begin{pmatrix} \mu \\ \mathbf{A}\mu + \mathbf{b} \end{pmatrix}, \quad \text{cov}[\mathbf{z}] = \begin{pmatrix} \mathbf{\Lambda}^{-1} & \mathbf{\Lambda}^{-1}\mathbf{A}^T \\ \mathbf{A}\mathbf{\Lambda}^{-1} & \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^T \end{pmatrix}.$$

(a) Show that the marginal distribution $p(\mathbf{x})$ is given by $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu, \mathbf{\Lambda}^{-1})$.

(b) Show that the conditional distribution $p(\mathbf{y}|\mathbf{x})$ is given by $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{Ax} + \mathbf{b}, \mathbf{L}^{-1})$.

Question 3

Show that the covariance matrix Σ that maximizes the log likelihood function is given by the sample covariance

$$\ln p(\mathbf{X}|\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^N(\mathbf{x}_n - \mu)^T \Sigma^{-1}(\mathbf{x}_n - \mu).$$

Is the final result symmetric and positive definite (provided the sample covariance is nonsingular)?

Hints

(a) To find the maximum likelihood solution for the covariance matrix of a multivariate Gaussian, we need to maximize the log likelihood function with respect to Σ . The log likelihood function is given by

$$\ln p(\mathbf{X}|\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^N(\mathbf{x}_n - \mu)^T \Sigma^{-1}(\mathbf{x}_n - \mu).$$

(b) The derivative of the inverse of a matrix can be expressed as

$$\frac{\partial}{\partial x}(\mathbf{A}^{-1}) = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\mathbf{A}^{-1}$$

We have the following properties

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr}(\mathbf{A}) = \mathbf{I}, \quad \frac{\partial}{\partial \mathbf{A}} \ln|\mathbf{A}| = (\mathbf{A}^{-1})^T.$$

Question 4

(a) Derive an expression for the sequential estimation of the variance of a univariate Gaussian distribution, by starting with the maximum likelihood expression

$$\sigma_{\text{ML}}^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2.$$

Verify that substituting the expression for a Gaussian distribution into the Robbins-Monro sequential estimation formula gives a result of the same form, and hence obtain an expression for the corresponding coefficients a_N .

(b) Derive an expression for the sequential estimation of the covariance of a multivariate Gaussian distribution, by starting with the maximum likelihood expression

$$\Sigma_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \mu_{\text{ML}})(\mathbf{x}_n - \mu_{\text{ML}})^T.$$

Verify that substituting the expression for a Gaussian distribution into the Robbins-Monro sequential estimation formula gives a result of the same form, and hence obtain an expression for the corresponding coefficients a_N .

Hints

(a) Consider the result $\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$ for the maximum likelihood estimator of the mean μ_{ML} , which we will denote by $\mu_{\text{ML}}^{(N)}$ when it is based on N observations. If we dissect out the contribution from the final data point \mathbf{x}_N , we obtain

$$\mu_{\text{ML}}^{(N)} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n = \frac{1}{N} \mathbf{x}_N + \frac{1}{N} \sum_{n=1}^{N-1} \mathbf{x}_n = \frac{1}{N} \mathbf{x}_N + \frac{N-1}{N} \mu_{\text{ML}}^{(N-1)}$$

(b) Robbins-Monro for maximum likelihood

$$\theta^{(N)} = \theta^{(N-1)} + a_{(N-1)} \frac{\partial}{\partial \theta^{(N-1)}} \ln p(x_N | \theta^{(N-1)}).$$

Question 5

Consider a D -dimensional Gaussian random variable \mathbf{x} with distribution $N(\mathbf{x}|\mu, \Sigma)$ in which the covariance Σ is known and for which we wish to infer the mean μ from a set of observations $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$. Given a prior distribution $p(\mu) = N(\mu|\mu_0, \Sigma_0)$, find the corresponding posterior distribution $p(\mu|\mathbf{X})$.

Program Question

Use online judge to solve this problem

In this coding exercise, you will implement the K-nearest Neighbors (KNN) algorithm. You are provided with a Jupyter Notebook **just for reference**. The requirement on online judge will be very different from the notebook.

This is a classification problem and we will use the Breast Cancer dataset:

| K | Norm | Accuracy(%) |
|---|-------|-------------|
| 3 | L1 | |
| 3 | L2 | |
| 3 | L-inf | |
| 5 | L1 | |
| 5 | L2 | |

| K | Norm | Accuracy(%) |
|---|-------|-------------|
| 5 | L-inf | |
| 7 | L1 | |
| 7 | L2 | |
| 7 | L-inf | |

Table1: Accuracy for the KNN classification problem on the validation set

A training data (X train) is provided which has several datapoints, and each datapoint is a p-dimensional vector (i.e., p features). Your task is to implement the K-nearest neighbors algorithm. Use the **Euclidean** distance.