CX 4803 CML – Spring 2022

Assignment 3

Please show your work as if you were explaining your solution to another student. In particular, please **show all your code***. used to generate the solutions. Submit your solutions as a single pdf file on Canvas. Include your program listings in your pdf file.*

1. Consider a multivariate Gaussian distribution with unknown mean μ, but known covariance matrix Σ. Given a prior distribution

$$p(\mu) = N(\mu|\mu_0, \Sigma_0),$$

and data $\mathbf{x}$, the posterior distribution is

$$p(\mu|\mathbf{x}) = N(\mu|\mu_n, \Sigma_n),$$

with

$$\mu_n = \Sigma_0 \left(\Sigma_0 + \frac{1}{n}\Sigma\right)^{-1} \mu_n + \frac{1}{n}\Sigma \left(\Sigma_0 + \frac{1}{n}\Sigma\right)^{-1} \mu_0$$

and

$$\Sigma_n = \Sigma_0 \left(\Sigma_0 + \frac{1}{n}\Sigma\right)^{-1} \frac{1}{n}\Sigma.$$

Suppose the multivariate Gaussian distribution has 3 dimensions and that

$$\Sigma = \begin{bmatrix} 1.0 & 0.3 & 0.0 \\ 0.3 & 0.9 & 0.2 \\ 0.0 & 0.2 & 0.8 \end{bmatrix}.$$

Choose μ_0 as the zero vector and $\Sigma_0 = I$ (the identity matrix).

(a) Write an expression that is the product of a likelihood function and a prior distribution in the form

$$p(\mathbf{x}|\mu)p(\mu) = c \cdot \exp(\ldots)$$

i.e., what is the expression inside the exponential function?

(b) Simulate 5 observations from $N(2I, \Sigma)$. Hint: use the Cholesky factorization. The value of 2 represents a value chosen by nature. Call these observations $x_1, \ldots, x_5$. Given this data, what is the posterior mean and covariance matrix?

(c) The posterior distribution could be treated as the prior distribution to compute a new posterior distribution after seeing another set of data. Suppose the new data that is observed is exactly the same set $x_1, \ldots, x_5$ that was observed earlier. What is the new posterior mean and covariance matrix?

(d) Now assume no data has been observed so that we have a prior distribution with $\mu_0 = 0$ and $\Sigma_0 = I$. Suppose you observe the 10 data values

$$x_1, \ldots, x_5, x_1, \ldots, x_5$$

that is, the data is repeated. What is the posterior mean and covariance matrix in this case?

(e) What observations can you make from the above experiment?

2. Consider polynomial regression using a degree 9 polynomial and a sample of 5 observations. To approximately solve $y \approx \Phi w$, the matrix Φ has fewer equations than columns, i.e., there are fewer equations than unknowns, so there might be an infinite number of solutions for w. In this case, how can we compute w?

   Now consider polynomial ridge regression. How does regularization deal with the above issue? Does the regression polynomial go through all the 5 observations?

3. Reproduce Figures 3.8 and 3.9 in the textbook by Bishop. Your figures are expected to be a little different due to the choice of parameters and random data. Be sure to state the parameter values that you used.