School of Electronics and Computer Science
University of Southampton

COMP6257(2022/23): Bayesian, Active and Reinforcement Learning Assignment One [30%]

Issue	17 Feb 2023
Deadline	10 Mar 2023
Feedback by	16 Mar 2023

5% The inverse of a partitioned matrix is given by:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix},$$

where $M = \left[A - BD^{-1}C\right]^{-1}$ (§2.3.1, PRML). Verify if the above is true. The inverse of a rank one update of a matrix is given by

$$\left[A + \boldsymbol{x}\boldsymbol{x}^T\right]^{-1} = A^{-1} - \frac{A^{-1}\boldsymbol{x}\boldsymbol{x}^T A^{-1}}{1 + \boldsymbol{x}^T A^{-1}\boldsymbol{x}}.$$

Verify if the above is true.

5% Bayesian estimation of the mean of a univariate Guassian, the variance of which is known is given in PRML §2.3.6 Eqn. 2.141 & 2.142. With reference to these two formulae briefly explain in your own words the effect of:

1. A very large number of observations, *i.e.* $N \to \infty$; and
2. A highly confident prior in the value of the mean.

10% Consider Bayesian analysis of the illustrative polynomial interpolation problem in PRML §1.1. using *Evidence Approximation* method discussed in PRML §3.5.1.

1. Generate a dataset similar to this illustration and solve the regression problem with various orders of polynomials. You should have a training dataset and a separate testing dataset. What is the variation in the test set prediction error taken over different realizations of the data? Quantify this by means of boxplots at different model orders.
2. Implement a regularized regression for a high order polynomial (say 9) at various values of the regularization hyperparameter λ.
3. Compute the Bayesian evidence formula of §3.5.1 for this problem at various values of α and β taken over a regular grid on two dimensions. Illustrate the results by means of contour plot.
4. By inspecting the above find the values of α and β for which the evidence $p(\boldsymbol{t}|\alpha, \beta)$ is maximum.
5. How close are these values to the solutions derived in Eqn. 3.92 and 3.95? How close is their ratio to the best λ obtained in (ii) above?

10% In a robot navigation problem, the decision to take a particular action is given by a logistic regression on two inputs $\boldsymbol{x} = [x_1,\ x_2]^T$. The robot has a programmed set of two weights $\boldsymbol{w} = [w_1,\ w_2]^T$ with uncertainties on them specified by a mixture Gaussian distribution:

$$p(\boldsymbol{w}) \;=\; \sum_{j=1}^{2} \mathcal{N}\left(\boldsymbol{w} | \boldsymbol{\mu}_j,\, \boldsymbol{\Sigma}_j\right),$$

where $\boldsymbol{\mu}_1 = [2.0,\ 5.0]^T$, $\boldsymbol{\mu}_1 = [3.0,\ 1.0]^T$, $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $\boldsymbol{\Sigma}_2 = 0.1\boldsymbol{I}$.

Given a particular input $[4.0,\ 4.0]^T$, the robot is to evaluate the probability of taking the action (and an uncertainty on it).

Use this contrived example to implement a (i) *Rejection Sampling*; and (ii) *Importance Sampling* approaches to derive the robot's probability of action. Compare the relative convergence performance of the two schemes.

Reference

PRML Bishop CM: *Pattern Recognition and Machine Learning*