CS 539 Machine Learning Homework 4

Conceptual and Theoretical Questions (3 questions, 20 pts)

1. Consider a hidden Markov model in which the emission densities are represented by a parametric model p(x|z,w), such as a linear regression model or a neural network, in which w is a vector of adaptive parameters. Describe how the parameters w can be learned from data using maximum likelihood. (This is question 13.4, from Bishop's textbook). **(8 pts)**

2. Show that the finite sample estimator f defined by (11.2) has mean equal to E[f] and variance given by (11.3) (This is question 11.1, from Bishop textbook). **(6 pts)**

3. Suppose that z has a uniform distribution over the interval [0, 1]. Show that the variable y = b tan z + c has a Cauchy distribution given by (11.16). (This is question 11.7, from Bishop's textbook). **(6 pts)**

Application Questions (4 questions, 80 pts)

Viterbi Decoding Algorithm (20 points) We discussed the Viterbi algorithm to find the most probable sequence of hidden states for a given observation sequence.

a. For the toy example explained in the attached pdf file (Example-Viterbi-DNA.pdf), write the Viterbi algorithm and repeat the result presented in the pdf file.
b. Repeat the decoding process in part (a) for the observed sequence of: AGTCGTA

Bayesian Filtering (20 points)
We covered the LDS in the class. A particular interest in LDS is the inference process, where we derive the posterior probability of x_k given the observation till time k – $y_{1...k}$. This is called filtering. In https://users.aalto.fi/~ssarkka/pub/cup_book_online_20131111.pdf, equation set (4.10-4.13), you will find the solution of the filter. This solution can be found in Bishop's textbook, section 13.3, as well.

Here, we build simulation data and derive inference step for a simple LDS model. Let's assume:

State model:

$$p(x_k|x_{k-1}) \sim N(0.99 * x_{k-1} + 0.1, 0.1)$$

Observation model:

$$p(y_k|x_k) \sim N(-2 * x_k + 1, 0.4)$$

and $p(x_0) \sim N(0,1)$.

a) Create a simulated data for x_0 to x_{100} and y_1 to y_{100} using the model described above.
b) Follow equation set (4.10-4.13) to estimate $p(x_k|y_{1...k})$ for $k = 1$ to 100. Note that you only observe y_1 to y_{100} and you build the posterior probability of x_k.
c) Show the posterior mean and its confidence overlayed on the ground truth x_k.
d) Repeat parts b and c for the observation provided in filter_problem.xlsx file. The first column is index, the second is x_k, and the last column is y_k.
e) Repeat part *d* using the following state process model.

$$p(x_k|x_{k-1}) \sim N(x_{k-1}, 0.2)$$

The observation model will be the same.
f) Discuss your results in *c*, *d*, and *e* parts.

Sequential MCMC (20 points)

In reading assignment 2, we learn about "particle filtering". Here, we use the SIS technique to solve the previous filtering problem.

a. Use 100 particles to estimate $p(x_k|y_{1\dots k})$ for filter_problem.xlsx file. Show the mean of your estimate and its confidence interval overlayed on the ground truth x_k. Use the state model discussed in part e of the previous problem.
b. Repeat the same procedure with 1000 particles.

GP and Linear Regression (20 points)

Gaussian Process is a powerful tool for regression and classification problems. Here, we will compare GP and linear regression prediction accuracy in a regression problem with the dataset presented in synchronous machine.csv file.

For this problem, you need to shuffle data and use 10-fold cross-validation.

a. Use GP with "exponentiated quadratic kernel" and show your test and training MSE. For the Kernel parameters, you might examine a set of different values and pick the one that provides the lowest MSE. Note that it is possible to estimate the kernel parameters as well.
b. Use linear regression using 4 predictors and show your test and training MSE.
c. Discuss the result in parts *a* and *b*

Dataset link: https://archive.ics.uci.edu/ml/datasets/Synchronous+Machine+Data+Set

GP Kernels: https://peterroelants.github.io/posts/gaussian-process-kernels/