Tuesday 26th June 2018
(2 hours)

ADVANCED STATISTICAL INFERENCE

Answer all questions.

This examination paper includes three questions and is worth a total of 80 marks.

1. Probabilistic Reasoning (total of [20] points)

(a) Consider three random variables A, B, and C. How many ways are there to express the joint distribution as a function of the marginals/conditionals?

[5]

(b) Write Bayes theorem to derive $p(A|B,C)$ from $p(B|A,C)$. How does this simplify if A does not depend on C?

[5]

(c) How can you obtain $p(A)$ from $p(A,B,C)$?

[5]

(d) Focus on A and B only, and imagine that these are binary variables and that you can obtain as many samples as you want from any of their joint/conditional/marginal distributions. How would you test whether A and B are independent?

[5]

2. Bayesian Linear Regression and Gaussian Processes (total of [35] points)

(a) What is Linear Regression and why is it called "Linear"?

[5]

(b) Explain how to treat Linear Regression in a Bayesian way.

[5]

(c) Explain how to use Linear Regression for the estimation of nonlinear functions.

[5]

(d) In Bayesian Linear Regression we usually assume that the labels are corrupted by noise. How would you reformulate Bayesian Linear Regression when assuming that the input too is affected by noise?

[5]

(e) What is the computational complexity (space and time) of Bayesian Linear Regression with respect to number of observations N and dimensionality of the features D

[5]

(f) Denote the $N \times D$ input matrix by X. Interpreting Gaussian Processes as Bayesian Linear Regression, what is the "equivalent" kernel of such Gaussian Processes?

[5]

(g) What is the computational complexity (space and time) of Gaussian Processes?

[5]

3. Supervised learning (total of [25] points)

(a) How would you go about extending the Naïve Bayes classifier to regression?

[10]

(b) What is a kernel function in Machine Learning? Why is it useful?

[5]

(c) Imagine a supervised learning problem where you apply a probabilistic model based on kernels. How would you optimize kernel parameters in the two cases where the kernel is parameterized by (i) one parameter or (ii) one-hundred parameters?

[5]

(d) Describe an example of a probabilistic model where we need to resort to approximate inference.

[5]