Monday 26th June 2017
(2 hours)

ADVANCED STATISTICAL INFERENCE

Answer all questions.

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

1. Imagine to observe some data $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ which are generated from an unknown process.

(a) Imagine making the assumption that these data come from a given density function with some parameters; write the likelihood function, explain any further assumptions on the independence of the data, and comment on the meaning of the likelihood.

[4]

(b) Describe how you would obtain the best set of parameters to model the given data.

[3]

(c) Describe the effect of having a prior that is zero everywhere except in the interval $[0, 1]$ on the posterior over model parameters.

[2]

(d) Describe a procedure to obtain the following predictive distribution: $p(\mathbf{x}_*|X)$.

[4]

(e) Imagine having a set of modeling assumptions $M_1, \ldots, M_k$ for the given data, with no preference over any of these; how would you pick the best model for the given data?

[4]

(f) Imagine again having a set of modeling assumptions $M_1, \ldots, M_k$ for the given data, with no preference over any of these; how would you use them all to obtain the predictive distribution $p(\mathbf{x}_*|X)$?

[3]

2. Gaussian processes

(a) What is a Gaussian process?

[6]

(b) Explaining the meaning of a kernel function in the context of Gaussian processes.

[6]

(c) What are the computational challenges associated with the use of Gaussian processes?

[4]

(d) Comment on the possibility to use Gaussian processes for classification tasks.

[4]

3. Supervised learning

(a) Describe the elements to construct a Bayesian classifier.

[4]

(b) What are the assumptions behind the Naïve Bayes classifier?

[2]

(c) Provide motivations and examples where the Naïve Bayes classifier may be preferable over the Bayesian classifier.

[2]

(d) Comment on the main differences between the Bayesian Classifier and a Bayesian version of logistic regression.

[4]

(e) What is the specificity of a classifier?

[4]

(f) What is the sensitivity of a classifier?

[4]

4. Projection

(a) Principal components analysis (PCA) is a popular projection technique. Describe what projection techniques do and give an application where projection would be useful.

[3]

(b) What property of the original data does PCA try to maintain in the projection?

[2]

(c) Briefly describe one way in which the number of components could be chosen.

[2]

(d) Imagine you were explaining PCA to someone with no knowledge of it. Draw a diagram to illustrate the operation of PCA that shows some hypothetical data and the direction of the first two principle components.

[3]

(e) A client has provided you with a classification dataset consisting of N observations, each of which consists of values for D variables, all of which are real valued. In addition, they have supplied an independent test set of T observations on which to evaluate any model you propose. You decide that you will use PCA as a pre-processing step to reduce the number of dimensions. Write a detailed plan detailing how you will perform the analysis. You may use any one (or more) suitable classification methods of your choice.

[10]

END OF QUESTION PAPER