EGT3
ENGINEERING TRIPOS PART IIB

Thursday 25 April 2024   2 to 3.40

Module 4M24

COMPUTATIONAL STATISTICS AND MACHINE LEARNING

Answer not more than ***three*** *questions.*

All questions carry the same number of marks.

The ***approximate*** *percentage of marks allocated to each part of a question is indicated in the right margin.*

Write your candidate number ***not*** *your name on the cover sheet.*

STATIONERY REQUIREMENTS
Write on single-sided paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

You may not remove any stationery from the Examination Room.

1 The Monte Carlo estimator is required for the integral of the function $f(x)$

$$\mathcal{I} = \int_a^b f(x)p(x)dx,$$

where $x \in [a, b]$ and $p(x)$ is the Lebesgue density. This estimator can be written as

$$\hat{\mathcal{I}} = \frac{1}{N}\sum_{n=1}^{N} f(x_n),$$

where each x_n is i.i.d. from $p(x)$.

(a) Assuming that $\mathbb{E}\{f(X)\} = \mathcal{I}$, and $\text{Var}\{f(X)\} = \sigma_f^2$, use the moment generating function to show that as $N \to \infty$, $\hat{\mathcal{I}}$ is normally distributed as

$$\hat{\mathcal{I}} \sim \mathcal{N}\left(\mathcal{I}, \frac{\sigma_f^2}{N}\right)$$

where the moment generating function for each $f(x)$ is $M_f(t) = \mathbb{E}\{\exp(tf(X))\}$. [60%]

(b) The generalised Monte Carlo estimator

$$\tilde{\mathcal{I}} = \sum_{n=1}^{N} w_n f(x_n)$$

is unbiased when $\sum_{n=1}^{N} w_n = 1$. What choice of weights $w_n, n = 1, \ldots, N$ gives the minimum variance unbiased estimator, i.e. when $\text{Var}\left\{\tilde{\mathcal{I}}\right\}$ is minimised? [40%]

2 The probability of data $y \in \mathbb{R}$ given model parameters θ is denoted by $p(y|\theta)$. Assuming a prior on θ as $p(\theta)$ the posterior follows as

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)},$$

where

$$p(y) = \int p(y|\theta)p(\theta)d\theta.$$

Consider the introduction of a tempering or annealing variable $t \in [0, 1]$ and $p(t)$ is uniform in $[0, 1]$. The tempered data probability is $p(y|\theta, t) \propto p(y|\theta)^t$.

(a) Give the expression for the tempered posterior $p(\theta|y, t)$. [10%]

(b) Give the expression for the conditional probability $p(y|t)$, and its value when $t = 0$ and $t = 1$, i.e. $p(y|t = 0), p(y|t = 1)$. [15%]

(c) Show that

$$\frac{d}{dt} \log p(y|t) = \mathbb{E}_{\theta|y,t} \{\log p(y|\theta)\},$$

where $\mathbb{E}_{\theta|y,t}\{\cdot\}$ denotes the expectation with respect to the tempered posterior $p(\theta|y, t)$. Hint: Note that x^t can be written in the form $x^t = \exp(t \log(x))$. [30%]

(d) Show that

$$\log p(y) = \mathbb{E}_{\theta,t|y} \{\log p(y|\theta)\},$$

where $\mathbb{E}_{\theta,t|y}\{\cdot\}$ denotes the expectation with respect to the tempered joint density $p(\theta, t|y) \propto p(\theta|y, t)p(t)$. [15%]

(e) Derive an MCMC algorithm that gives a Monte Carlo estimate of $\log p(y)$. [30%]

3 The Lebesgue measure for the intervals $[a_d, b_d]$ in $d = 1, \ldots D$ takes the form

$$\prod_{d=1}^{D} (b_d - a_d)\,.$$

Note that the Lebesgue measure is finite, monotonic and translation invariant.

(a) Show that in a complete infinite dimensional Hilbert space, the Lebesgue measure is not well defined. [35%]

(b) The standard Gaussian measure in $\mathbb{R}$ is given as

$$g(B) = \frac{1}{\sqrt{2\pi}} \int_B \exp\left(-\frac{x^2}{2}\right) dx.$$

Consider $\mathbb{R}^\infty$ and the product measure $\mu = \prod_{k=1}^{\infty} g_k$, with each g_k equivalent to g. Under what conditions will this product measure be well defined? What subspace of $\mathbb{R}^\infty$ do these conditions represent? [15%]

(c) The covariance of the Gaussian measure μ can be considered as an Identity operator. A more general linear covariance operator C on the Hilbert space $\mathcal{H}$ can also be defined.

(i) Give conditions on C which need to be satisfied for the infinite dimensional measure $\mu_C = \mathcal{N}(0, C)$ to be well defined. [20%]

(ii) Show that the trace operator of C in Hilbert space $\mathcal{H}$, having an orthonormal basis $e_n, n = 1, \ldots, \infty$, defined on a domain Ω can be written as

$$\sum_{n=1}^{\infty} \int_\Omega \int_\Omega e_n(x) c(x, y) e_n(y) dx dy,$$

where $c(x, y)$ is the covariance function corresponding to the operator C. [30%]

4 Consider two univariate Gaussian probability densities on $x \in \mathbb{R}$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(-\frac{(x-\mu_f)^2}{2\sigma_f^2}\right), \quad g(x) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp\left(-\frac{(x-\mu_g)^2}{2\sigma_g^2}\right).$$

(a) Show that the product of the two densities takes the form of a scaled (unnormalised) Gaussian $p_{fg}(x) \propto f(x)g(x)$ with mean and variance μ_{fg}, σ_{fg}^2 respectively. Give the expressions for μ_{fg} and σ_{fg}^2. [40%]

(b) Derive the form of the normalised density

$$p_{fg}(x) = \frac{f(x)g(x)}{\int_{\mathbb{R}} f(y)g(y)dy}.$$

[20%]

(c) Derive an Unadjusted Langevin Algorithm (ULA) to sample from $p_{fg}(x)$ for the case where $\mu_f = 0$, $\mu_g = 1$ and $\sigma_f = \sigma_g = 1$. [40%]

END OF PAPER

THIS PAGE IS BLANK