

UNIVERSITY OF YORK

MLAP

MACHINE LEARNING AND APPLICATIONS

---

# Open Examination

---

*Examination number:*

Y3606797

# Contents

<b>1</b>	<b>Task 1</b>	<b>2</b>
1.1	What sort of model? . . . . .	2
1.2	EM in general . . . . .	2
1.3	EM for the current model . . . . .	2

# 1 Task 1

## 1.1 What sort of model?

The naive Bayes model makes the assumption that all features are conditionally independent given the class. The given Bayesian network explicitly models the conditional independence in its structure.

## 1.2 EM in general

The EM algorithm is an iterative approach to maximising the likelihood when we have hidden variables. We have a model  $p(v, h|\theta)$  and we wish to find  $\theta$  by maximising the marginal likelihood  $p(v|\theta)$ . EM achieves this by replacing the marginal likelihood with a lower bound, the bound depends on  $\theta$  and the set of variational distributions  $\{q\}$ , we optimise this bound. This is done by first fixing  $\theta$  and then optimising *w.r.t*  $\{q\}$ , then fixing  $\{q\}$  and optimising *w.r.t*  $\theta$ . These are known as the 'E' and 'M' steps and are repeated until convergence. The EM algorithm may not always return the correct MLEs because it can get stuck on local optima.

## 1.3 EM for the current model

Firstly we assume some initial parameters for the BN, so we have  $\theta_{X1}^0$ ,  $\theta_{X2}^0$ ,  $\theta_{X3}^0$ ,  $\theta_H^0$  and  $\theta_C^0$ . Then we compute the distribution for the hidden data H:

$$q_{t=1}^{n=1}(H) = p(H|C = 0, X1 = 0, X2 = 0, X3 = 1, \theta^0),$$

$$q_{t=1}^{n=2}(H) = p(H|C = 1, X1 = 1, X2 = 0, X3 = 0, \theta^0)...$$

and so on for each data point in the training data. This is first E-step in the algorithm.

The first M-step involves maximising the energy term by choosing new values for  $\theta$ . The energy term is given by the following equation:

$$E(\theta) = \sum_{n=1}^N \{ \langle \log p(C^n|H^n) \rangle_{q_t^n(H)} \langle \log p(X1^n|H^n) \rangle_{q_t^n(H)} + \langle \log p(X2^n|H^n) \rangle_{q_t^n(H)} \\ + \langle \log p(X3^n|H^n) \rangle_{q_t^n(H)} + \langle \log p(H^n) \rangle_{q_t^n(H)} \} \quad (1)$$