



MSc Examination

Friday 8th May 2014 14:30 - 17:00

ECS708 Machine Learning Duration: 2 hours 30 minutes

**YOU ARE NOT PERMITTED TO READ THE CONTENTS OF THIS QUESTION PAPER UNTIL
INSTRUCTED TO DO SO BY AN INVIGILATOR**

Answer FOUR questions

If you answer more questions than specified, only the first answers (up to the specified number) will be marked. Cross out any answers that you do not wish to be marked

Calculators are not permitted in this examination. Please state on your answer book the name and type of machine used.

Complete all rough workings in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately. It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

EXAM PAPERS MUST NOT BE REMOVED FROM THE EXAM ROOM

Examiners: Dr. Ioannis Patras, Dr. Chrisantha Fernando

(2014)

© Queen Mary, University of London, 2014

Question 1

- a) Define the joint probability $p(A, B)$ and the conditional probability $p(A|B)$. You may want to use a diagram/sketch. Give the formula that relates them.

[4 marks]**Answer:**

The conditional probability $p(A|B)$ is the probability that the event A occurs, given that the event B has occurred. The joint probability is the probability that both events occurred. Their relation is given by the formula $p(A|B) = P(A, B) / P(B)$

- b) Give the relation between the joint probability $P(X, Y)$ and the probabilities $P(X)$ and $P(Y)$ that holds in the case that X and Y are independent random variables. Give the condition that holds when X and Y are uncorrelated. Are these conditions the same?

[4 Marks]**Answer:**

X, and Y are independent if $p(x, y) = p(x)p(y)$. They are uncorrelated if their covariance is zero, that is if $\text{var}\{x, y\} = E\{(x - E\{x\})(y - E\{y\})\} = E\{xy\} - E\{x\}E\{y\} = 0$. The condition of independence is much stronger.

- c) Show that the expected value of the sum of two independent random variables X and Y is equal to the sum of the expected values of X and Y. That is, show that $E\{X + Y\} = E\{X\} + E\{Y\}$. You may show it either for continuous or discrete variables. (Hint: You need to work with the joint probability $P(X, Y)$).

[6 Marks]**Answer**

$$\begin{aligned}
& E\{X+Y\} \\
&= \int_{X,Y} (X+Y)p(X,Y)dxdy = \int_{X,Y} (X+Y)p_{XY}(X,Y)dxdy \\
&= \int_{X,Y} Xp_{XY}(X,Y)dxdy + \int_{X,Y} Yp_{XY}(X,Y)dxdy \\
&= \int_{X,Y} Xp_X(X)p_Y(Y)dxdy + \int_{X,Y} Yp_X(X)p_Y(Y)dxdy \\
&= \int_{X,Y} Xp_X(X)p_Y(Y)dxdy + \int_{X,Y} Yp_X(X)p_Y(Y)dxdy \\
&= \int_X \int_Y Xp_X(X)p_Y(Y)dxdy + \int_X \int_Y Yp_X(X)p_Y(Y)dxdy \\
&= \int_X Xp_X(X)dx \int_Y p_Y(Y)dy + \int_X p_X(X)dx \int_Y Yp_Y(Y)dy \\
&= \int_X Xp_X(X)dx + \int_Y Yp_Y(Y)dy = E\{X\} + E\{Y\}
\end{aligned}$$

- d) An IT worker works from home 2 days a week. When she works from home there is a 30% chance she will NOT answer an email within an hour, 10% chance that she will not answer an email within two hours, and it is certain that she will answer all the emails within the day. When she is at office, there is a 50% chance she will NOT answer an email within one hour, 10% chance that she will not answer an email within the two hours, and it is certain that she will answer all the emails within the day.
- If you send her an email, what is the probability that she will answer within 2 hours?
 - Given that she hasn't replied to your email within 1 hour, what is the probability that she is working from home? Does the information that she hasn't answered the email within 1 hour makes it more or less likely that she works from home?

[11 Marks]

Answer

The joint probability table is given by:

$$\begin{array}{ccc}
0.4*0.7 & 0.4*0.2 & 0.4*0.1 \\
0.6*0.5 & 0.6*0.4 & 0.6*0.1
\end{array} = \begin{array}{ccc} .28 & 0.08 & 0.04 \\ .3 & .24 & .06 \end{array}$$

The probability that she will answer within two hours is the sum of the values of the first two columns and equals 0.90

Let D denote the event that she will answer after one hour. $P(D) = 0.42$ (sum of last two columns)

(2014)

Let H denote the event that she works at home. $P(H) = 0.4$ (sum of first row)

$P(D,H) = 0.12$ is the probability that she works from home AND will answer after one hour.

$$P(H|D) = P(D,H) / P(D) = 0.12 / 0.42 = 0.2857 < p(H) = 0.4$$

That means that if she has not answered the email within the first hour, this gives some evidence that she is at work.

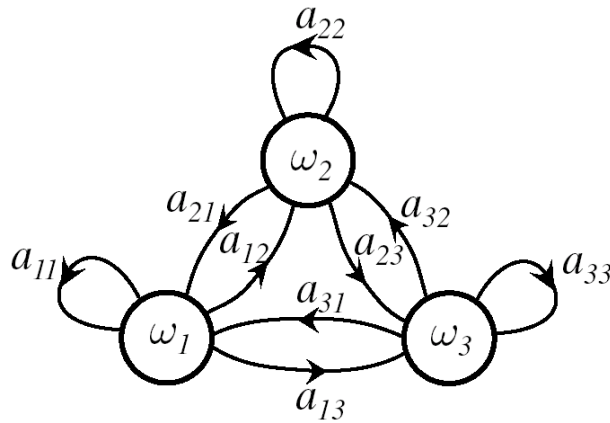
[Total 25 marks]

Question 4

- a) With a help of a diagram explain the main principles of a first-order Markov Model. Explain what is meant by the term ‘‘first-order’’. What are the differences with a hidden Markov model (HMM)? In your answer, define the states ω_i , the symbols v_k , and the matrices $\mathbf{A}=[a_{ij}]$ and $\mathbf{B}=[b_{jk}]$.

[6 marks]

Answer:



A first-order Markov model is a generalisation of a finite state machine, in the sense that transitions between the states are probabilistic. In the example above there are three states ω_i , and transitions probabilities a_{ij} from state i to state j . It is called a first-order model meaning that the probability of going to state j from state i depends only on the current state i and not on states further in the past.

In a hidden Markov model:

At each state j an observation symbol v_k is emitted with probability b_{jk}

We do not have direct evidence of the state, we observe only the symbols.

- b) The decoding problem can be stated as follows: Given an HMM and a sequence of observation symbols $\mathbf{V}^{1:T}$ determine the most likely sequence of hidden states $\omega^{1:T}$. What are the other two types of problems that arise in the context of HMMs?

[6 marks]

Answer:

(2014)

- 1) The evaluation problem: Given an HMM (that is, $\omega_i(0)$, $\mathbf{A}=[a_{ij}]$ and $\mathbf{B}=[b_{jk}]$), what is the probability of generating the sequence $\mathbf{V}^{1:T}$.
- 2) The Learning problem: Given the structure of the HMM, that is the number of states, the connections and the symbols) learn the state transition matrix and the symbol emission matrix. These are learned from a set of observation sequences.
- c) Describe an algorithm that solves the decoding problem, as this is described in part (b). What is the name of this algorithm?

[13 marks]

Answer:

The most likely sequence of hidden states can be estimated by the Viterbi algorithm, as given below.

1. Initialize: $\delta_i(0) = 1$ for init state $\omega(0)$, 0 otherwise; $t \leftarrow 0$
2. Recursion: $t \leftarrow t + 1$
3. $\delta_j(t) \leftarrow \max_i [\delta_i(t-1)a_{ij}] b_{j,v(t)}$
4. $\psi_j(t) \leftarrow \arg \max_i [\delta_i(t-1)a_{ij}]$
5. If $t < T$ repeat from 2.
6. Termination: $\delta(T) \leftarrow \arg \max_i \delta_i(T)$
7. Backtracking:
8. $t \leftarrow t - 1$; $\omega(t) \leftarrow \psi_{\omega(t+1)}(t+1)$
9. If $t > 1$ repeat from 8.
10. Output state sequence $\{\omega(0), \dots, \omega(T)\}$.