UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE	:	**COMP0137**
ASSESSMENT PATTERN	:	**COMP0137A7UC / COMP0137A7PC**
MODULE NAME	:	**COMP0137 - Machine Vision**
LEVEL:	:	**Undergraduate (Masters Level) / Postgraduate**
DATE	:	**11/05/2021**
TIME	:	**10:00**

This paper is suitable for candidates who attended classes for this module in the following academic year(s):

Year
2020/21

Additional material	N/A
Special instructions	N/A
Exam paper word count	N/A

TURN OVER

UCL Computer Science Examination paper

Paper details

Academic year:	2020/21
Module title:	Machine Vision
Module code:	COMP0137
Exam period:	Main summer examination period
Duration:	24 hours
Deliveries for which intended:	A7P (taught postgraduate, level 7) A7U (undergraduate, level 7)
Cohorts for which intended:	2020/21; 2019/20

Instructions

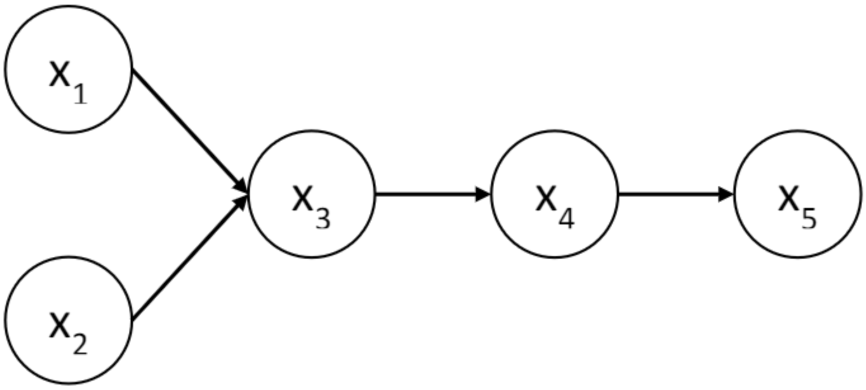
You must answer ALL questions on ALL pages of this assignment.

A maximum of 100 marks is available. The marks available for each part of each question are indicated in square brackets [n].

Submit your answers online through Moodle. Unless specified, you do not need to show your work.

Section A: A Graphical Model

1. There is a Markov Model pictured below. Each discrete variable is illustrated to show what values are possible. A "configuration" is a specific combination, where each x_i is in a specific state, for i=1..5. The unary costs, pairwise costs, and triplet costs for each variable or combination of variables are specified in the tables provided.



All variables are discreet, but they do not all have the same number of possible states. For example, variable X_1 can be “A” or “B”, while X_2 can be “C”, “D”, or “E”. The unary costs for all five variables are:

$U_1(w_1 = A)$	2.0	$U_2(w_2 = C)$	2.1	$U_3(w_3 = F)$	1.3	$U_4(w_4 = I)$	0.9	$U_5(w_5 = L)$	2.0
$U_1(w_1 = B)$	2.9	$U_2(w_2 = D)$	0.1	$U_3(w_3 = G)$	2.7	$U_4(w_4 = J)$	1.1	$U_5(w_5 = M)$	3.1
-	-	$U_2(w_2 = E)$	0.9	$U_3(w_3 = H)$	1.2	$U_4(w_4 = K)$	1.7	$U_5(w_5 = N)$	1.8

The pairwise and triplet costs are here:

$P_4(w_3, w_4)$

	$w_4 = I$	$w_4 = J$	$w_4 = K$
$w_3 = F$	1.0	1.5	2.0
$w_3 = G$	1.5	1.0	1.5
$w_3 = H$	2.0	1.5	1.0

$P_5(w_4, w_5)$

	$w_5 = L$	$w_5 = M$	$w_5 = N$
$w_4 = I$	1.0	1.5	2.0
$w_4 = J$	1.5	1.0	1.5
$w_4 = K$	2.0	1.5	1.0

$T_3(w_3 = F, w_1, w_2)$

	$w_2 = C$	$w_2 = D$	$w_2 = E$
$w_1 = A$	1.0	2.0	2.5
$w_1 = B$	2.0	3.0	3.5

$T_3(w_3 = G, w_1, w_2)$

	$w_2 = C$	$w_2 = D$	$w_2 = E$
$w_1 = A$	2.0	2.0	4.0
$w_1 = B$	2.0	1.0	2.0

$T_3(w_3 = H, w_1, w_2)$

	$w_2 = C$	$w_2 = D$	$w_2 = E$
$w_1 = A$	3.5	4.0	3.5
$w_1 = B$	4.0	2.0	1.0

Use the above information to answer the following questions.

(a) Configuration A is ACGJN. Compute the total cost of this configuration: _____

[7 marks]

(b) Configuration B is AEFKM. Compute the total cost of this configuration: _____

[7 marks]

(c) What is the Maximum Likelihood configuration? _ _ _ _ _

[7 marks]

(d) What is the Maximum a Posteriori configuration? _ _ _ _ _

[8 marks]

[Total for Question 1: 29 marks]

Section B: Solar Energy Task

2. Imagine your company develops solar-energy farms, for electricity generation. Every year, you consider multiple empty fields throughout the UK, as possible sites to install your simple solar panels, which are mounted flat, parallel to the ground. Business, weather, and environmental factors drive you to build such collection sites very selectively: you want to be fairly certain a field will have a specific energy-yield before building an appropriately-sized set of solar panels there.

Once panels are installed, it is easy to measure their minute-by-minute power generation. But for practical reasons, assume it is very hard to install panels prospectively. Instead, assume it is easy to install a wide-field-of-view dome camera, to collect sky-images over long periods.

In your first year, you have three or four existing solar-energy farms, where you can both measure the electricity yield, and snap corresponding dome-camera photos. This is your initial source of training data. You will be training and applying a compute vision model that infers energy yield from dome-camera images.

[Question cont. over page]

(a) Even without any energy-yield information, it's possible to train an auto-encoder using the dome-camera images. Name one benefit of doing this.

Answer: ______

[6 marks]

(b) You can collect and analyze the paired data, so energy yield and corresponding dome-camera images, every month. Why might it be unwise to make decisions just from the data collected in January, the first month of operations?

Answer: ___________

[6 marks]

(c) You could train the model that predicts energy-yield using data in the dual domain. Under what circumstances would this approach be preferable?

Answer: ___________

[6 marks]

(d) You could train a Maximum A Posteriori (MAP) or a Bayesian model. Describe one example scenario where the difference between the two matters.

Answer: ___________

[6 marks]

(e) The potential sites, where initial filming using the dome-cameras has happened, are laid out geographically on a map. The layout of these sites is distinctly not grid-like. How would this affect the pairwise potentials, if you use a Markov Random Field to regularize your energy-yield predictions?

Answer: ___________

[6 marks]

[Total for Question 2: 30 marks]

Section C: Miscellaneous Questions

3. There are pairs of probability distributions, where the second is conjugate to the first. Now consider a situation where the second distribution IS NOT conjugate to the first. Give one sentence for each, to describe a) one potential benefit of this design choice to use non-conjugate-distributions, and b) one potential disadvantage.

____

____

____

____

[6 marks]

4. We wish to determine the pose of a crash-test dummy, from a single input image. There is exactly one dummy visible in the photo, and we'll assume it is not occluded by anything else. We are given a simplified initialization algorithm, where a limb-detector runs on the given input image. The detector detects a loose bounding box around each separate limb of a crash-test dummy, e.g. a "limb" is detected for the chest, left upper arm, the left lower arm, the left hand, etc. However, the detector does not distinguish between these limbs, so each is just one limb in an unordered list of detections.

(a) Using a maximum of two sentences, explain how would you practically obtain the training data needed to determine the pairwise terms for a tree-model. For context, you would use those terms to find the Maximum A Posteriori configuration of limb-labels for all the visible parts of the crash test dummy, replacing each generic "limb" label by its inferred body-part name.

____

____

[6 marks]

[Questions continued on next page]

(b) Now, instead of each label w_i representing just the limb's body-part name, you want to determine each limb's 3D position and 3D orientation. Using a maximum of two sentences, describe why you might regret choosing dynamic programming to find the MAP solution to this harder version of the problem? Use a 3rd sentence to name an alternative and say what this alternative would sacrifice compared to dynamic programming.

____

____

____

[6 marks]

[Total for Question 4: 12 marks]

5. Find out what an "Event Camera" is. You may access Wikipedia, or any published writing or manuals online. For each of the scenarios below, indicate whether it's better to feed data to a computer vision algorithm from (choose one): a) A regular CMOS camera, b) An Event camera, or c) Both cameras are equally effective in that scenario.

5.a: Detect an animal sneaking around a garden at night. _________

5.b: Generate a deblurred colour wedding photo. _________

5.c: Feed a face recognition system for keyless entry to a secure building. _________

5.d: Inspect train tracks from a moving train. _________

5.e: Detect forgery in paper banknotes. _________

[12 marks total]

6. Give two distinct reasons why lines that are parallel to each other in the real world can appear non-parallel in a photograph.

____

[4 marks]

[Questions continued on next page]

7. Your aim is to buy the right number of roof tiles, to replace an existing roof. Imagine you record a colour image sequence while walking all the way around the outside of a home. The house is standing in an open clearing, so no trees or other objects block your view as you point the camera at it. You use the image sequence to first compute a sparse 3D point cloud. Then you reconstruct a dense 3D point cloud, and proceed to segment which 3D points belong to the roof. In one sentence, describe the one critical step that is still needed to compute the surface area of the existing roof.

____

____

[4 marks]

[Total for Question 3, 4, 5, 6, and 7: 41 marks]

[Total for all questions: 100 marks]