MV simulation exam

Question 1 (2’)

State the extent of independence (if any) between the variables, if Pr(x1,x2) = Pr(x1)Pr(x2).

Feedback

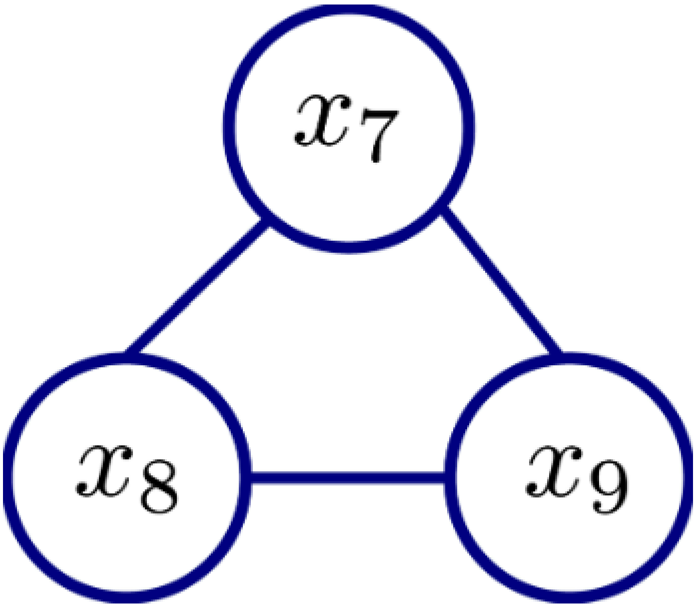
x1 and x2 are independent of each other.

Reasoning: Short-answer questions are marked by hand, so don't worry if the Moodle auto-grading reports that you're getting a question on this practice-exam wrong. Auto-grading is used on the questions which are multiple-choice (so in effect, 99% of questions).

This question is worth only 2 points (i.e. "marks") out of 100. Typical questions are worth 4 or 5 points. So questions worth fewer points should be easier: a concise correct answer is enough, and/or it's obvious, and probably doesn't take much creative effort. Similarly, questions worth more points are harder. So beware of true/false questions worth 10 points! (That will never happen, but a 4-pointer IS reasonable and would be about 2x as hard as this example question.)

The correct answer is: x1 and x2 are independent

Question 2 (2’)



State the extent of independence (if any) between the variables in the model pictured here:

Feedback

None.
None of the variables is independent, nor are they conditionally independent.

The correct answer is: no independence

Question 3 (6’)

In maximum 2 sentences, explain one way to make Earth mover's distance more tolerant to outliers.

Feedback

Model answer: penalize distance up to a point, but then use a truncated quadratic (for example) so that distances beyond a parameter d all cost the same.

Reasoning: 6 marks because this topic was not explicitly covered. It requires that you go find and read a few paragraphs on your own on a trusted site to understand Earth Mover's Distance**. Then it'll help to remember how we handled outliers in class (Gamma distribution was just one approach). The distance between clusters typically grows linearly, which makes sense, since it should cost more to move "dirt" longer distances. To keep that behavior but also minimize the damage done by a far-away outlier, you could instead measure the distance between cluster i and cluster j using a specially selected "robust" function, like truncated linear, truncated quadratic, or similar.

Sources: Wikipedia is on the "good" side of the boundry, when you're deciding what sources to use. As a tertiary source, it's not perfect (course materials are best and use the same terminology consistently). But Wikipedia cites its sources and is curated by a surprising number of people who pay close attention to accuracy. I often check its accuracy when wording an exam question, so that students who use it aren't at a disadvantage (though the article may have many distractor topics that are hard to parse in a short time). HOWEVER using an AI during the exam is just as disallowed (against the rules!) for this class as communicating with another person. I know that the temptation is huge - maybe you'll get away with it, while improving your grade somewhat? That is cheating. In contrast to how I check Wikipedia, I also check the main AI engines when writing exam questions - and I favor questions where the AI's responses are misleading, or even better - where the AI gets it wrong. That's not always possible, and I'm sorry if this feels adversarial just to counter-act a tiny handful of students who may be tempted, but it beats the alternative of having closed-book exams in a 150-person exam hall.

Why ask questions about a topic we didn't cover? Why this topic? Several reasons. Working during the term and studying the prescribed material should lead to marks above-50%, while top marks have to be a little more challenging. Plus, with an open-book exam, memorization isn't worth testing - but application of principles is. The field of computer vision is vast and continues to grow; EMD is also referred to as the Wasserstein Metric, which you're likely to encounter if you work with GANs at some point.

Question 4 (5’)

You are trying to train a neural network **Θ** on dataset **D** of satellite images with corresponding labels. The label for each image is a binary mask indicating the presence of water, e.g. lakes, rivers, and even small tributaries. However, while training, your computer reports an out-of-memory error. Among possible actions you can take now, the least-good option is to:

a. Reduce the batch size

b. Get a computer with more memory

c. Reduce the image width and height by 2x.

d. Use gradient accumulation

e. Reduce the number of layers in the network

Feedback

Correct choice: Reduce the image width and height by 2x.

For d, A reasonable strategy, so not the outlier-answer we're looking for. It's a tougher answer to eliminate though, since we didn't cover this in class, so this would require a quick search on Google or similar (not-AI) search engines. Also, it's not always effective, and could be more work for less benefit than reducing the batch size - but it's not wrong.

For e, Not the wrong answer we're seeking: shrinking the NN's size helps it fit in memory, even though it could hurt the model's ability to learn the correct relationship between inputs and outputs.

Reasoning: This is the best answer because this is the least-appropriate solution. It would mean that smaller rivers would disappear from the input RGB's and from the ground-truth label masks.

Note about choice "Use gradient accumulation": A reasonable strategy, so not the outlier-answer we're looking for. It's a tougher answer to eliminate though, since we didn't cover this in class, so this would require some reading. Also, it's not always effective, and could be more work for less benefit than reducing the batch size - but it's not wrong.

The correct answer is:

Reduce the image width and height by 2x.

Question 5 (2’)

Two texture-rich images are related to each other by a specific homography **H**, but **H** is unknown.

How many interest-point pairs, at a minimum, are needed to compute **H**?

Feedback

4

Reasoning: Need four distinct pairs of points, where for each of the four, you know the 2D coordinates in each image corresponding to a single real-world 3D point.

The correct answer is: 4

Question 6 (3’)

In one sentence, describe a scenario where even having n corresponding pairs of points across two images will NOT yield a valid homography **H**?

Feedback

It won't work if any three of the points are co-linear.

Reasoning: This is less of a memory test. Rather, you'd need to remind yourself how a homography is computed, and think about when matrix algebra can give degenerate or unexpected results. That happens most often around matrix inversion, and when a matrix is rank-deficient, leaving the problem under-determined.

The correct answer is: co-linear

Question 7 (6’)

Name two user-determined parameters that you'll need to set to run the RANSAC algorithm when finding a good homography **H**.

Feedback

The number of iterations, and the inlier-threshold.

Reasoning: Not discussed but also acceptable: clock-time threshold, though this is uncommon. Written answers are checked by hand, so the wording and spelling don't need to be a perfect match.

Such a question will typically be turned into multiple-choice, so that students who studied the course materials can get easy'ish points, and can spend more of their time on other questions. In multiple-choice form, the wrong answers will also sound credible to someone who doesn't know the algorithm, e.g. the number of points, the number of layers, the SVD of H, the rank-deficiency score, and other possibly made-up terms.

The correct answer is: iterations tolerance

Question 8 (8’)

You are asked to design and train a computer vision model to do greenscreening, i.e. a kind of chromakeying. The model should predict the alpha channel **A**, with the same width and height as the input RGB image.

In a max of four sentences total, describe two benefits of using synthetic training data, being specific or giving examples as part of your answer.

Feedback

Once you set up synthetic data generation, it'll be easy to generate infinite quantities of paired input/output training images/alpha channels.
(This part of the answer is quite obvious)
You can generate training data with exquisite ground-truth labels (alpha channels), which would be hard to get from real images. For example, filming a real object that is semi-transparent (e.g. translucent material or just the object-boundary) would produce an input RGB, for which even a human expert could struggle to label a per-pixel alpha value.
You can prepare custom training data for various situations: both the foreground and background that you expect on a particular shoot can be used to train a bespoke model. For example, use people to train a model for dialog shots, and plastic models for Star-Wars space ships.

(only the first two benefits offered in the answer will be marked)

The correct answer is: Talk about quantity of training data, covering the variety of situations, including hard-to-get ground truth like transparency.

Question 9 (6’)

You are still designing and training an alpha channel regressor.

In one sentence, describe one disadvantage of using synthetic training data that comes from simple rendering of 3D models in Blender.

Then, in a second sentence, describe something you can do with that synthetic data (so NOT using real data directly) to mitigate that disadvantage. Give an example or be specific to get full marks.

Feedback

This question has two parts, for 3 and 3 points respectively (though you wouldn't know that as a test-taker).
With a question worth 6 points, the reader should surmize that each of two sentences has room for a "great" answer, beyond an obvious-to-everyone generic one.

First, a generic answer would say something about the synthetic data being unrepresentative of real data.
A better answer would say that synthetic data could be biased, with all the background pixels being the same shade of chromakey-green, or all the foreground objects being imaged without motion blur, or without any mpeg- or jpeg-compression. Any of those would lead to overfitting, especially since the validation set would be made up of the same type of data (so any iterative model training would proceed to overfit).

Second, a generic answer would say "data augmentation." That's barley worth a point, as it's pretty obvious.
A better answer would explain that knowing something about real-world special effects could help design algorithms to "corrupt" the synthetic data to more closely resemble real-world filming situations, e.g. uneven lighting (spatially) or changing lighting on the greenscreen over time such as from clouds, or compression artifacts from digital cameras. If the first part of the answer talked about motion blur, then the second one could talk about how like-real motion blur could be simulated with the synthetic data after-the-fact, with different velocities, to match the intended test-cases.

The correct answer is: bias and corrupt