CV2 SS2023

1. **Probabilistic Models**
   1. How is the posterior distribution of a computer vision task related to the observation model? Explain the observation model in a typical probabilistic approach to stereos.
2. *In computer vision tasks like stereo vision, the posterior distribution represents the probability of model parameters (e.g., depth) given observed data. The observation model specifies the likelihood of observing data given these parameters.*
3. *In stereoscopy, it describes how likely it is to see a pair of stereo images based on the disparity or depth map. Bayes' theorem combines this likelihood with a prior distribution, yielding the posterior distribution that updates our estimate of parameters based on observed data and prior knowledge.*

*(add probabilistic computation here)*

* 1. What is the reason to use Gaussian distributions in modeling a posterior despite their sensitivity to outliers?

*Gaussian distributions are favored for modeling a posterior due to their* ***analytical simplicity****,* ***convenience in Bayesian inference****, and* ***adherence to the Central Limit Theorem****. Besides, their* ***computational efficiency*** *and* ***widespread applicability*** *make them a popular choice.*

* 1. Briefly explain the concept of model robustness and how we can measure it. Provide an example of robust likelihood-function.

*Model robustness reflects a model's ability to* ***maintain performance******amidst diverse conditions****. It involves assessing resilience to adversarial examples, generalization across domains, noise tolerance, and consistent predictions. A robust likelihood function, exemplified by* ***Huber loss****, minimizes sensitivity to outliers in Bayesian statistics. The Huber loss combines squared and absolute errors, offering a balance that reduces the impact of extreme values on parameter estimation, making it more robust in the presence of noisy data.*

* 1. If X and Y are **random variables** and E[Y|X]=constant, then show that X and Y are uncorrelated. (Hint: It is sufficient to show that Cov(X,Y)=E[XY]-E[X]E[Y]=0. )

random variables **=>** X, Y are independent => **E[XY|X]** = E[Y|X]E[X] [1]

E[Y|X] = constant = k => E[XY] = E[**E[XY|X]**] [2]

[1]+[2] => E[XY] = E[E[Y|X]E[X]] = E[Y|X]E[X]

=> Cov(X,Y) = E[XY]-E[X]E[Y] = E[Y|X]E[X]−E[X]E[Y]

= k⋅ E[X]−E[X]E[Y] = k⋅ E[X]− k⋅ E[X] = 0

* 1. What are the advantages and disadvantages of using probabilistic models in computer vision over non-probabilistic methods?

***Advantages****:*

*Probabilistic models offer uncertainty quantification, essential for handling real-world variability. They enable principled integration of prior knowledge and adaptability to diverse data conditions. Additionally, probabilistic models naturally handle missing or noisy data.*

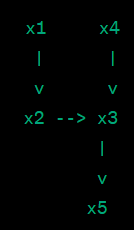
***Disadvantages****:*

*Probabilistic models can be computationally demanding due to the need for inference algorithms. Model complexity and interpretability might suffer, impacting real-time applications. Handling uncertainty can be challenging, and model performance heavily depends on the quality of prior information and assumptions, which may introduce biases.*

1. Graphical Models
   1. Why are Directed Graphical Models(DGMs) useful?

*DGMs are valuable for modeling and analyzing complex systems due to their ability to represent causal relationships. With directed edges indicating the direction of influence, DGMs offer a clear depiction of cause-and-effect dependencies among variables. This facilitates efficient inference, aiding predictions and decision-making under uncertainty. The graphical structure enhances computational efficiency, enabling the estimation of probability distributions and making DGMs practical tools for understanding and navigating real-world scenarios where variables interact in a directed and often causal manner.*

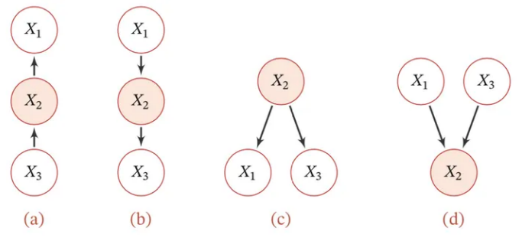
* 1. We have a probability distribution over the variables x1, x2, x3, x4, x5 and the following independence statements:

Draw a directed graphical model with at least four edges from which (at least) these independence statements can be inferred.

*1. edge from x1 to x2 indicates independence: .*

*2. edge from x4 to x3, conditioned on x2, implies independence: .*

*3. edge from x1 to x5, conditioned on x3, implies independence: .*

**

*(brief explanation of DGM: )*

1. *, when x2 is known.*
2. *, when x2 is known.*
   1. What are the two basic types of graphical models that were presented in the lecture? Why is it useful to distinguish between those? Name one (additional) property in how they differ.
   2. Explain in your own words the basic idea behind using the Potts model as a compatibility function in stereo.
   3. A model describing the relationship between three random events A, B, and C is given by p(A, B, C) = p(C | A, B) p(A) p(B). Is A independent of B? Is A conditionally independent of B given C?
3. Optimization
   1. We have an energy function over a vector of binary labels

With the following definition of the energy terms:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x1 | U1(x1) | | | x2 | U2(x2) | x3 | | U3(x3) |
| 0 | 3 | | | 0 | 3 | 0 | | 1 |
| 1 | 1 | | | 1 | 2 | 1 | | 4 |
| xi | | xi+1 | Pi,i+1(xi,xi+1) | | | |
| 0 | | 0 | 0 | | | |
| 0 | | 1 | 2 | | | |
| 1 | | 0 | 2 | | | |
| 1 | | 1 | 0 | | | |

Draw the graph for which we can use graph cuts to compare the lowest energy via the minimum cut and annotate each edge with its weight. Name the general requirement on the pairwise energy terms in order to solve a binary labeling problem such as this one to global optimality using graph cuts.

* 1. List two advantages of computing marginal distributions over the MAP estimate. Name two algorithms used to compute the marginals. Explain if it is possible for two distributions to have the same MAP estimate, but different marginal distributions?
  2. When using probabilistic modeling for computing the disparity map, we can make use of continuous optimization methods. In this case, we need to use interpolation. Explain when and why?
  3. Name one major issue of doing message passing in Markov Random Fields with loops instead of in a tree graph.
  4. What does loopy belief propagation compute?

1. Image Restoration
   1. Name two classical and two modern noise removal techniques.
   2. We typically assume that the noise is conditionally independent, additive, and Gaussian distributed when modeling the possibility by image denoising. While this assumption leads to our model working in many applications, it can be sub-optimal in other cases. Name two such scenarios with examples.
   3. Name a common issue when using linear filters for noise removal. Explain how bilateral filters try to solve this issue.
   4. Assume that we want to model properties of natural images with a pairwise MRE. Name a property of natural images that is better represented by Student-t potentials than Gaussian potentials. Name a property of natural images that we can not model with a pairwise MRF, regardless of the used potentials.
   5. You have an image with a lot of self-similarity in which you want to reduce noise. Name a method presented in the lecture that is particularly well suited for this scenario and explain why.
   6. Using Bayes’ rule, write the equation for image super resolution as probabilistic inference. Assuming a zoom factor of 2, what is a simple likelihood model for image super resolution?
2. Optical Flow / Stereo
   1. Explain two different ways of how CNNs can be applied to the disparity estimation task from stereo images.
   2. Briefly explain the aperture problem in the context of optical flow estimation. What consequences does it have for optical flow estimation?
   3. If we consider stereo matching for large images, we have to model a likelihood term over several thousands or even millions of pixels. Which assumption did we use in class to make the likelihood tractable? Why is such an assumption reasonable?
   4. What are multi-scale loss function? Why are they beneficial for training optical flow networks?
   5. Why does the method of Lukas-Kanade only work for small motions?
   6. Name two principles from traditional optical flow estimation that are incorporated in PWC-Net.
   7. List a few reasons for poor generalization of deep-learning-based networks for flow estimation on unseen data.
3. Tracking
   1. Define tracking and explain how it is distinct from optical flow in its definition.
   2. What is a rigid object, articulated object, and non-rigid object in the context of tracking? Give an example for each one of them.
   3. What are two main advantages of representing the posterior in tracking with a discrete set of samples or particles?
   4. Which restrictions does a Kalman filter have? Name a setting in which these restrictions might failure in tracking.
   5. Name an advantage of using Kalman filters and an advantage of using Particle filtering.
   6. Why does tracking balls in a billiard game become more difficult with a higher number of balls?
4. Segmentation
   1. Compute the energy values of a Potts model.

with

For the given segmentation below, where the numbers denote the label of a node xi in the graphical model. (δ denotes the set of all edges) If you are unsure about your result, explain how you computed it.

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 1 | 2 | 3 |
| 1 | 1 | 2 | 3 |
| 2 | 2 | 3 | 3 |

* 1. What is the intuition behind using a contrast-sensitive Potts model for segmentation?
  2. How can you decide whether a given segmentation for an image is a good one?
  3. Explain the SLIC algorithm in detail.
  4. When creating superpixels for segmentation, we often make use of affinities like color and intensity for encoding the similarity between pixels. What are the drawbacks of using (i) color affinity and (ii) intensity affinity when creating superpixels?
  5. Explain the steps necessary to convert a backbone network for image classification into a deep network for semantic image segmentation. Why do we also need to add skip connections and what do they help with?