Graphical Models
Discrete Inference and Learning
Lecture 1

MVA
2020 – 2021

http://thoth.inrialpes.fr/~alahari/disinflearn

Slides based on material from Stephen Gould, Pushmeet Kohli, Nikos Komodakis, M. Pawan Kumar, Carsten Rother, Daphne Koller, Dhruv Batra
Graphical Models ?
What this class is about?

- Making **global** predictions from **local** observations
  
  **Inference**

- Learning such models from large quantities of data
  
  **Learning**
Motivation

• Consider the example of medical diagnosis

Slide inspired by PGM course, Daphne Koller
Motivation

- A very different example: image segmentation

Millions of pixels
Colours / features

Pixel labels
{building, grass, cow, sky}

e.g., [He et al., 2004; Shotton et al., 2006; Gould et al., 2009]

Slide inspired by PGM course, Daphne Koller
Motivation

• What do these two problems have in common?

Slide inspired by PGM course, Daphne Koller
Motivation

• What do these two problems have in common?
  – Many variables
  – Uncertainty about the correct answer

Graphical Models (or Probabilistic Graphical Models) provide a framework to address these problems
(Probabilistic) Graphical Models

• First, it is a model: a declarative representation
• Can also define the model
  – with domain knowledge
  – from data
(Probabilistic) Graphical Models

• Why probabilistic?
• To model uncertainty
• Uncertainty due to:
  – Partial knowledge of state of the world
  – Noisy observations
  – Phenomena not observed by the model
  – Inherent stochasticity
(Probabilistic) Graphical Models

• Probability theory provides
  – Standalone representation with clear semantics
  – Reasoning patterns (conditioning, decision making)
  – Learning methods
(Probabilistic) Graphical Models

• Why graphical?
• Intersection of ideas from probability theory and computer science
  – To represent large number of variables

Predisposing factors
Symptoms
Test results

Random variables \( Y_1, Y_2, \ldots, Y_n \)

Goal: capture uncertainty through joint distribution \( P(Y_1, \ldots, Y_n) \)

Slide inspired by PGM course, Daphne Koller
(Probabilistic) Graphical Models
(Probabilistic) Graphical Model

• Examples

Bayesian network (directed graph)

Markov network (undirected graph)

Figure courtesy: D. Koller
(Probabilistic) Graphical Model

• Examples

Diagnosis network: Pradhan et al., UAI’94

Segmentation network (Courtesy D. Koller)
(Probabilistic) Graphical Model

• Intuitive & compact data structure

• Efficient reasoning through general-purpose algorithms

• Sparse parameterization
  – Through expert knowledge, or
  – Learning from data

Slide inspired by PGM course, Daphne Koller
(Probabilistic) Graphical Model

• Many many applications
  – Medical diagnosis
  – Fault diagnosis
  – Natural language processing
  – Traffic analysis
  – Social network models
  – Message decoding
  – Computer vision: segmentation, 3D, pose estimation
  – Speech recognition
  – Robot localization & mapping

Slide courtesy: PGM course, Daphne Koller
Image segmentation

Image

No graphical model

With graphical model

Sturgess et al., 2009
Multi-sensor integration: Traffic

• Learn from historical data to make predictions

Slide courtesy: Eric Horvitz, MSR
Stock market
Going global: Local ambiguity

- Text recognition

Smyth et al., 1994
Going global: Local ambiguity

• Textual information extraction

e.g., Mrs. Green spoke today in New York. Green chairs the financial committee.
Overview of the course

• Representation
  – How do we store $P(Y_1, \ldots, Y_n)$
  – Directed and undirected (model implications/assumptions)

• Inference
  – Answer questions with the model
  – Exact and approximate (marginal/most probable estimate)

• Learning
  – What model is right for data
  – Parameters and structure
First, a recap of basics
Graphs

• Concepts
  – Definition of G
  – Vertices/Nodes
  – Edges
  – Directed vs Undirected
  – Neighbours vs Parent/Child
  – Degree vs In/Out degree
  – Walk vs Path vs Cycle
Special graphs

- Trees: undirected graph, no cycles
- Spanning tree: Same set of vertices, but subset of edges, connected and no cycles

Slide courtesy: D. Batra
Directed acyclic graphs (DAGs)

Figure courtesy: D. Batra
Interpreting Probability

• What does $P(A)$ mean?

• Frequentist view
  – Limit $N \to \infty$, $(A \text{ is true})/N$
  – i.e., limiting frequency of a repeating non-deterministic event

• Bayesian view
  – $P(A)$ is your belief about $A$

Slide courtesy: D. Batra
Joint distribution

• 3 variables
  – Intelligence (I)
  – Difficulty (D)
  – Grade (G)

• Independent parameters?

Example courtesy: PGM course, Daphne Koller
Conditioning

• Condition on $g^1$

<table>
<thead>
<tr>
<th>I</th>
<th>D</th>
<th>G</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.126</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^2$</td>
<td>0.168</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^3$</td>
<td>0.126</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.009</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^1$</td>
<td>$g^2$</td>
<td>0.045</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^1$</td>
<td>$g^3$</td>
<td>0.126</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.252</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^2$</td>
<td>0.0224</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^3$</td>
<td>0.0056</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.06</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^2$</td>
<td>0.036</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^3$</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Example courtesy: PGM course, Daphne Koller
Conditioning

• $P(Y = y \mid X = x)$
• Informally,
  – What do you believe about $Y=y$ when I tell you $X=x$?

• $P(\text{France wins a football tournament in 2021})$?
• What if I tell you:
  – France won the world cup 2018
  – Hasn’t had catastrophic results since 😊
Conditioning: Reduction

- Condition on $g^1$

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>D</th>
<th>G</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.126</td>
<td></td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.252</td>
<td></td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

Example courtesy: PGM course, Daphne Koller
### Conditioning: Renormalization

<table>
<thead>
<tr>
<th>I</th>
<th>D</th>
<th>G</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>i₀</td>
<td>d₀</td>
<td>g₁</td>
<td>0.126</td>
</tr>
<tr>
<td>i₀</td>
<td>d₁</td>
<td>g₁</td>
<td>0.009</td>
</tr>
<tr>
<td>i₁</td>
<td>d₀</td>
<td>g₁</td>
<td>0.252</td>
</tr>
<tr>
<td>i₁</td>
<td>d₁</td>
<td>g₁</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Unnormalized measure

<table>
<thead>
<tr>
<th>I</th>
<th>D</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>i₀</td>
<td>d₀</td>
<td>0.282</td>
</tr>
<tr>
<td>i₀</td>
<td>d₁</td>
<td>0.02</td>
</tr>
<tr>
<td>i₁</td>
<td>d₀</td>
<td>0.564</td>
</tr>
<tr>
<td>i₁</td>
<td>d₁</td>
<td>0.134</td>
</tr>
</tbody>
</table>

Example courtesy: PGM course, Daphne Koller
Conditional probability distribution

• Example $P(G \mid I, D)$

<table>
<thead>
<tr>
<th></th>
<th>$g^1$</th>
<th>$g^2$</th>
<th>$g^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i^0,d^0$</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>$i^0,d^1$</td>
<td>0.05</td>
<td>0.25</td>
<td>0.7</td>
</tr>
<tr>
<td>$i^1,d^0$</td>
<td>0.9</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>$i^1,d^1$</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Example courtesy: PGM course, Daphne Koller
Conditional probability distribution

\[ p(x, y \mid Z = z) = \frac{p(x, y, z)}{p(z)} \]

Slide courtesy: Erik Sudderth
Marginalization

\[ P(I,D) \]

\[ \text{Marginalize I} \]

Example courtesy: PGM course, Daphne Koller
Marginalization

• Events
  – \( P(A) = P(A \text{ and } B) + P(A \text{ and } \text{not } B) \)

• Random variables
  – \( P(X = x) = \sum_{y} P(X = x, Y = y) \)
Marginalization

\[ p(x, y) = \sum_{z \in \mathcal{Z}} p(x, y, z) \]

\[ p(x) = \sum_{y \in \mathcal{Y}} p(x, y) \]

Slide courtesy: Erik Sudderth
Factors

• A factor $\Phi(Y_1,\ldots,Y_k)$

  $\Phi: \text{Val}(Y_1,\ldots,Y_k) \rightarrow \mathbb{R}$

• Scope = $\{Y_1,\ldots,Y_k\}$
Factors

• Example: $P(D, I, G)$

<table>
<thead>
<tr>
<th>I</th>
<th>D</th>
<th>G</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.126</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^2$</td>
<td>0.168</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^3$</td>
<td>0.126</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.009</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^1$</td>
<td>$g^2$</td>
<td>0.045</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^1$</td>
<td>$g^3$</td>
<td>0.126</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.252</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^2$</td>
<td>0.0224</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^3$</td>
<td>0.0056</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.06</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^2$</td>
<td>0.036</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^3$</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Example courtesy: PGM course, Daphne Koller
Factors

• Example: $P(D, I, g^1)$

<table>
<thead>
<tr>
<th></th>
<th>$D$</th>
<th>$G$</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i^0$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.126</td>
</tr>
<tr>
<td>$i^0$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.009</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^0$</td>
<td>$g^1$</td>
<td>0.252</td>
</tr>
<tr>
<td>$i^1$</td>
<td>$d^1$</td>
<td>$g^1$</td>
<td>0.06</td>
</tr>
</tbody>
</table>

What is the scope here?

Example courtesy: PGM course, Daphne Koller
General factors

- Not necessarily for probabilities

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a^0</td>
<td>b^0</td>
<td>30</td>
</tr>
<tr>
<td>a^0</td>
<td>b^1</td>
<td>5</td>
</tr>
<tr>
<td>a^1</td>
<td>b^0</td>
<td>1</td>
</tr>
<tr>
<td>a^1</td>
<td>b^1</td>
<td>10</td>
</tr>
</tbody>
</table>

Example courtesy: PGM course, Daphne Koller
Factor product

Example courtesy: PGM course, Daphne Koller
Factor marginalization

Example courtesy: PGM course, Daphne Koller
### Factor reduction

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a^1</td>
<td>b^1</td>
<td>c^1</td>
<td>0.25</td>
</tr>
<tr>
<td>a^1</td>
<td>b^1</td>
<td>c^2</td>
<td>0.35</td>
</tr>
<tr>
<td>a^1</td>
<td>b^2</td>
<td>c^1</td>
<td>0.08</td>
</tr>
<tr>
<td>a^1</td>
<td>b^2</td>
<td>c^2</td>
<td>0.16</td>
</tr>
<tr>
<td>a^2</td>
<td>b^1</td>
<td>c^1</td>
<td>0.05</td>
</tr>
<tr>
<td>a^2</td>
<td>b^1</td>
<td>c^2</td>
<td>0.07</td>
</tr>
<tr>
<td>a^2</td>
<td>b^2</td>
<td>c^1</td>
<td>0</td>
</tr>
<tr>
<td>a^2</td>
<td>b^2</td>
<td>c^2</td>
<td>0</td>
</tr>
<tr>
<td>a^3</td>
<td>b^1</td>
<td>c^1</td>
<td>0.15</td>
</tr>
<tr>
<td>a^3</td>
<td>b^1</td>
<td>c^2</td>
<td>0.21</td>
</tr>
<tr>
<td>a^3</td>
<td>b^2</td>
<td>c^1</td>
<td>0.09</td>
</tr>
<tr>
<td>a^3</td>
<td>b^2</td>
<td>c^2</td>
<td>0.18</td>
</tr>
</tbody>
</table>

>{|   |   |   |   |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a^1</td>
<td>b^1</td>
<td>c^1</td>
<td>0.25</td>
</tr>
<tr>
<td>a^1</td>
<td>b^2</td>
<td>c^1</td>
<td>0.08</td>
</tr>
<tr>
<td>a^2</td>
<td>b^1</td>
<td>c^1</td>
<td>0.05</td>
</tr>
<tr>
<td>a^2</td>
<td>b^2</td>
<td>c^1</td>
<td>0</td>
</tr>
<tr>
<td>a^3</td>
<td>b^1</td>
<td>c^1</td>
<td>0.15</td>
</tr>
</tbody>
</table>
| a^3 | b^2 | c^1 | 0.09 |}
Why factors?

• Building blocks for defining distributions in high-dimensional spaces

• Set of basic operations for manipulating these distributions
Independent random variables

\[ P(x, y) = p(x)p(y) \]

for all \( x \in \mathcal{X}, y \in \mathcal{Y} \)

Slide courtesy: Erik Sudderth
Marginal independence

- **Sets** of variables $X$, $Y$

- $X$ is independent of $Y$
  - Shorthand: $P 	riangleright (X \perp Y)$

- **Proposition**: $P$ satisfies $(X \perp Y)$ if and only if
  - $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$, $\forall x \in \text{Val}(X), y \in \text{Val}(Y)$
Conditional independence

• **Sets** of variables $X$, $Y$, $Z$

• $X$ is independent of $Y$ given $Z$ if
  – Shorthand: $P \vdash (X \perp Y \mid Z)$
  – For $P \vdash (X \perp Y \mid \emptyset)$, write $P \vdash (X \perp Y)$

• **Proposition:** $P$ satisfies $(X \perp Y \mid Z)$ if and only if
  – $P(X,Y \mid Z) = P(X \mid Z) \cdot P(Y \mid Z)$, $\forall x \in \text{Val}(X), y \in \text{Val}(Y), z \in \text{Val}(Z)$
Bayes Rule

• Simple yet profound

• Concepts
  – Likelihood
    • How much does a certain hypothesis explain the data?
  – Prior
    • What do you believe before seeing any data?
  – Posterior
    • What do we believe after seeing the data?
Bayesian Networks

• DAGs
  – nodes represent variables in the Bayesian sense
  – edges represent conditional dependencies

• Example
  – Suppose that we know the following:
    • The flu causes sinus inflammation
    • Allergies cause sinus inflammation
    • Sinus inflammation causes a runny nose
    • Sinus inflammation causes headaches
  – How are these connected?

Slide courtesy: Dhruv Batra
Bayesian Networks

• Example

Slide courtesy: Dhruv Batra
Bayesian Networks

• A general Bayes net
  – Set of random variables
  – DAG: encodes independence assumptions
  – Conditional probability trees
  – Joint distribution

\[
P(Y_1, \ldots, Y_n) = \prod_{i=1}^{n} P(Y_i | Pa_{Y_i})
\]
Bayesian Networks

• A general Bayes net
  – How many parameters?
    • Discrete variables $Y_1, \ldots, Y_n$
    • Graph: Defines parents of $Y_i$, i.e., $(Pa_{Y_i})$
    • CPTs: $P(Y_i | Pa_{Y_i})$

Slide courtesy: Dhruv Batra
Markov nets

• Set of random variables

• Undirected graph
  – Encodes independence assumptions

• Factors

Comparison to Bayesian Nets?
Pairwise MRFs

• Composed of pairwise factors
  – A function of two variables
  – Can also have unary terms

• Example
Markov Nets: Computing probabilities

• Can only compute ratio of probabilities directly

• Need to normalize with a **partition function**
  – Hard! (sum over all possible assignments)

• In Bayesian Nets, can do by multiplying CPTs
Markov nets $\iff$ Factorization

- Given an undirected graph $H$ over variables $Y = \{Y_1, ..., Y_n\}$
- A distribution $P$ factorizes over $H$ if there exist
  - Subsets of variables $S^i \subseteq Y$ s.t. $S^i$ are fully-connected in $H$
  - Non-negative potentials (factors) $\Phi_1(S^1), ..., \Phi_m(S^m)$: clique potentials
  - Such that
    \[
    P(Y_1, ..., Y_n) = \frac{1}{Z} \prod_{i=1}^{m} \Phi_i(S^i)
    \]
Conditional Markov Random Fields

- Also known as: Markov networks, undirected graphical models, MRFs
- Note: Not making a distinction between CRFs and MRFs
- $\mathbf{X} \in \mathcal{X}$: observed random variables
- $\mathbf{Y} = (Y_1, \ldots, Y_n) \in \mathcal{Y}$: output random variables
- $\mathbf{Y}_c$ are subset of variables for clique $c \subseteq \{1, \ldots, n\}$
- Define a factored probability distribution

$$P(\mathbf{Y} \mid \mathbf{X}) = \frac{1}{Z(\mathbf{X})} \prod_c \psi_c(\mathbf{Y}_c; \mathbf{X})$$

Partition function $= \sum_{\mathbf{Y} \in \mathcal{Y}} \prod_c \psi_c(\mathbf{Y}_c; \mathbf{X})$  

Exponential number of configurations!
MRFs / CRFs

• Several applications, e.g., computer vision

Low-level vision problems

Interactive figure-ground segmentation [Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006]

Surface context [Hoiem et al., 2005]

Semantic labeling [He et al., 2004; Shotton et al., 2006; Gould et al., 2009]

Stereo matching [Kolmogorov and Zabih, 2001; Scharstein and Szeliski, 2002]

Image denoising [Felzenszwalb and Huttenlocher 2004]
MRFs / CRFs

• Several applications, e.g., computer vision
MRFs / CRFs

• Several applications, e.g., medical imaging
MRFs / CRFs

• Inherent in all these problems are graphical models

Pixel labeling
Object detection
Pose estimation
Scene understanding
Maximum a posteriori (MAP) inference

\[ y^* = \arg\max_{y \in \mathcal{Y}} P(y \mid x) \]

\[ = \arg\max_{y \in \mathcal{Y}} \frac{1}{Z(x)} \prod_c \psi_c(y_c; x) \]

\[ = \arg\max_{y \in \mathcal{Y}} \log \left( \frac{1}{Z(x)} \prod_c \psi_c(y_c; x) \right) \]

\[ = \arg\max_{y \in \mathcal{Y}} \sum_c \log \psi_c(y_c; x) - \log Z(x) \]

\[ = \arg\max_{y \in \mathcal{Y}} \sum_c \log \psi_c(y_c; x) - E(\mathbf{Y}; \mathbf{X}) \]
Maximum a posteriori (MAP) inference

$$y^* = \arg\max_{y \in \mathcal{Y}} P(y \mid x) = \arg\max_{y \in \mathcal{Y}} \sum_c \log \Psi_c(Y_c; X)$$

$$= \arg\min_{y \in \mathcal{Y}} E(y; x)$$

MAP inference $\Leftrightarrow$ Energy minimization

The energy function is $E(Y; X) = \sum_c \psi_c(Y_c; X)$

where $\psi_c(\cdot) = -\log \Psi_c(\cdot)$
Clique potentials

• Defines a mapping from an assignment of random variables to a real number

\[ \psi_c : \mathcal{V}_c \times \mathcal{X} \rightarrow \mathbb{R} \]

• Encodes a preference for assignments to the random variables (lower is better)

• Parameterized as

\[ \psi_c(y_c; x) = w^T_c \phi_c(y_c; x) \]

Parameters
Clique potentials

- Arity

\[ E(y; x) = \sum_{c} \psi_c(y_c; x) \]

\[ = \sum_{i \in \mathcal{V}} \psi_i^U(y_i; x) + \sum_{ij \in \mathcal{E}} \psi_{ij}^P(y_i, y_j; x) + \sum_{c \in \mathcal{C}} \psi_c^H(y_c; x). \]

- Unary
- Pairwise
- Higher-order
Clique potentials

• Arity

4-connected, $\mathcal{N}_4$

8-connected, $\mathcal{N}_8$
Reason 1: Texture modelling

Training images

Test image

Test image (60% Noise)

Result MRF
4-connected
(neighbours)

Result MRF
4-connected

Result MRF
9-connected
(7 attractive; 2 repulsive)
Reason 2: Discretization artefacts

Higher-connectivity can model true Euclidean length

[Boykov et al. ’03; ’05]
Graphical representation

• Example

\[ E(y) = \psi(y_1, y_2) + \psi(y_2, y_3) + \psi(y_3, y_4) + \psi(y_4, y_1) \]
Graphical representation

• Example

\[ E(y) = \sum_{i,j} \psi(y_i, y_j) \]
Graphical representation

• Example

\[ E(\mathbf{y}) = \psi(y_1, y_2, y_3, y_4) \]
A Computer Vision Application

Binary Image Segmentation

How?

Cost function Models our knowledge about natural images

Optimize cost function to obtain the segmentation
A Computer Vision Application

Binary Image Segmentation

Object - white, Background - green/grey

Each vertex corresponds to a pixel

Edges define a 4-neighbourhood grid graph

Assign a label to each vertex from $L = \{\text{obj}, \text{bkg}\}$

Graph $G = (V,E)$
A Computer Vision Application

Binary Image Segmentation

Object - white, Background - green/grey

Cost of a labelling $f : V \rightarrow L$

Cost of label ‘obj’ low  Cost of label ‘bkg’ high

Graph $G = (V,E)$

Per Vertex Cost
A Computer Vision Application

Binary Image Segmentation

Object - white, Background - green/grey

Cost of a labelling $f : V \rightarrow L$

Cost of label ‘obj’ high Cost of label ‘bkg’ low

UNARY COST

Graph $G = (V,E)$

Per Vertex Cost
A Computer Vision Application

Binary Image Segmentation

Object - white, Background - green/grey

Cost of a labelling $f : V \rightarrow L$

Cost of same label low

Cost of different labels high

Graph $G = (V,E)$

Per Edge Cost
A Computer Vision Application

Binary Image Segmentation

Object - white, Background - green/grey

Graph $G = (V,E)$

Cost of a labelling $f : V \rightarrow L$

Cost of same label high

Per Edge Cost

Cost of different labels low

PAIRWISE COST
A Computer Vision Application

Binary Image Segmentation

Object - white, Background - green/grey

Graph $G = (V, E)$

Problem: Find the labelling with minimum cost $f^*$
A Computer Vision Application

Binary Image Segmentation

Graph $G = (V,E)$

Problem: Find the labelling with minimum cost $f^*$
Another Computer Vision Application

Stereo Correspondence

How?

Minimizing a cost function
Another Computer Vision Application

Stereo Correspondence

Graph $G = (V,E)$

Vertex corresponds to a pixel

Edges define grid graph

$L = \{\text{disparities}\}$
Another Computer Vision Application

Stereo Correspondence

Cost of labelling $f$:

Unary cost + Pairwise Cost

Find minimum cost $f^*$
The General Problem

Graph \( G = (V, E) \)

Discrete label set \( L = \{1, 2, \ldots, h\} \)

Assign a label to each vertex \( f: V \to L \)

Cost of a labelling \( Q(f) \)

Unary Cost

Pairwise Cost

Find \( f^* = \arg \min Q(f) \)
Overview

• Basics: problem formulation
  – Energy Function
  – MAP Estimation
  – Computing min-marginals
  – Reparameterization

• Solutions
  – Belief Propagation and related methods [Lecture 3]
  – Graph cuts [Lecture 2]