Grundlagen der stochastischen Bildanalyse

1 Einführung

Zwei Beispiele zur Motivation

Abbildung 1: Bild der berühmten Lena

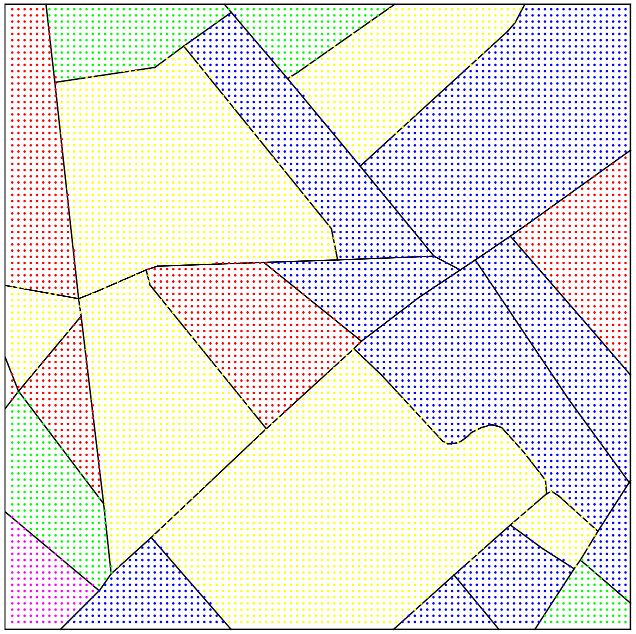

Abbildung 2: Bilddaten von Pariser Strassenzellen

- In den verschiedensten Gebieten treten Daten in Form von Bildern auf unterschiedlichste Weise auf.
- Grössenordnungen von makroskopisch über mesoskopisch und mikroskopisch bis zu nanoskopisch.
- Gebiete: Medizin, Informatik, Biologie, Mathematik,...
- Bilder basieren in der Regel auf einzelnen Pixeln (picture element).
- Verschieden Arten: Binär-, Graustufen-, Farbbilder

Bildarten

- Binärbild: Jedes Pixel nimmt den Wert 0 oder 1 an (entspricht Weiss und Schwarz).
- Graustufenbild: Jedes Pixel nimmt einen Wert zwischen 0 und K an (z.B. $K = 255$).
- Farbbild: Jedes Pixel besitzt drei Komponenten, die jeweils Werte zwischen 0 und K_i annehmen. (3 Farben Rot, Blau und Grün).

Weshalb wird Bildverarbeitung/Bildanalyse benötigt ?

- Wiederherstellung des Originalbilds da dieses (gewollt oder ungewollt) verändert (gestört) wurde.
- Quantitative und Qualitative Analyse des Bilds (z.B. Detektion von Kanten)
- Unterscheidung in stochastische und deterministische Ansätze.

2 Markovsche zufällige Felder

Zufällige Felder

Sei $\mathcal{I}$ die Menge der Pixel eines Bildes, wobei $i \in \mathcal{I}$ der Koordinate eines bestimmten Pixels entspricht.

Dann bildet eine Familie von Zufallsvariablen $\{X_i : i \in \mathcal{I}\}$ mit $X_i : \Omega \to \mathcal{X}$ ein *zufälliges Feld*.

Definiere $X_A = \{X_i : i \in A\}$ und $X_{-A} = \{X_i : i \in \mathcal{I} \backslash \mathcal{A}\}$.

Nachbarschaftsbeziehungen

- Zwischen den einzelnen Pixeln wird i.a. eine Nachbarschaftsrelation $\sim$ definiert.
- $\sim$ ist symmetrisch, d.h. falls $i \sim j$ dann $j \sim i$.
- Per Definition sei i kein Nachbar von sich selbst.
- Sei δ_i die Menge der Nachbarn von i.
- Eine Teilmenge von Pixeln bei der alle Pixel Nachbarn zueinander sind wird als *Clique* bezeichnet.
- Es können verschiedene Nachbarschaftsbeziehungen verwendet werden, vgl. Abb. 3.

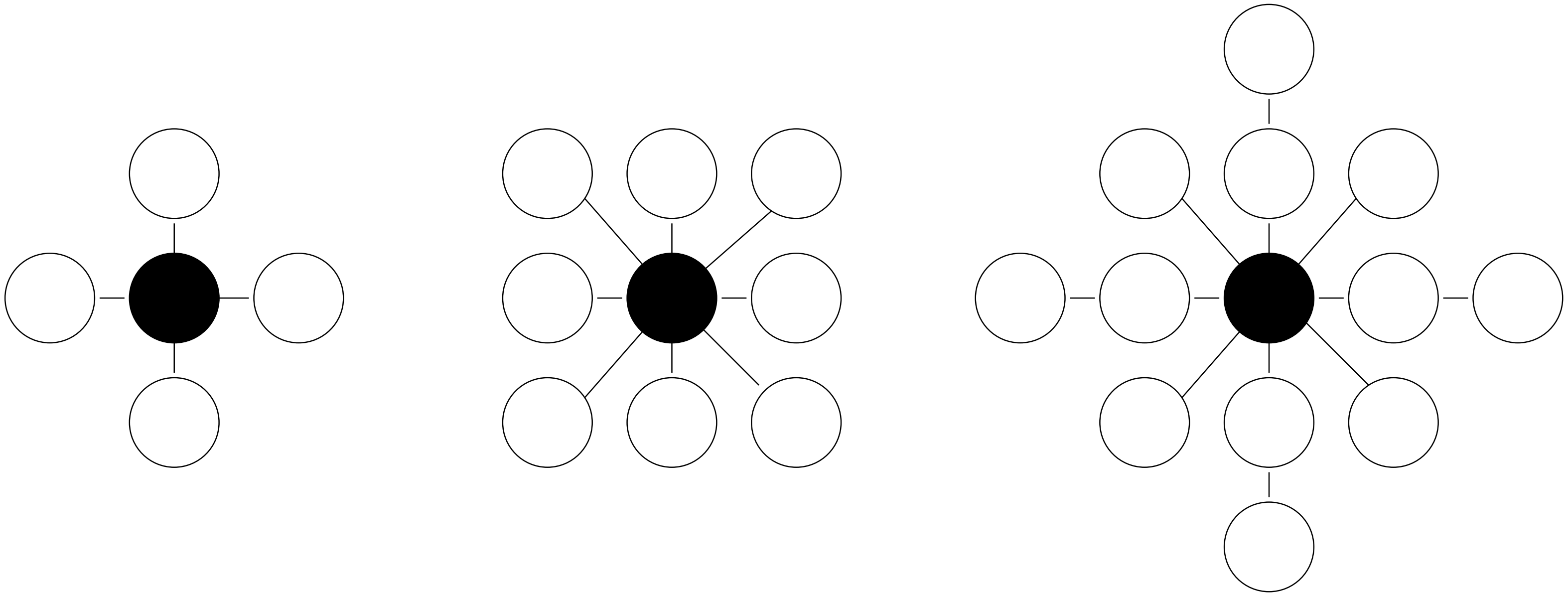

Abbildung 3: Vierer-, Achter- und Zwölfer Nachbarschaftsstrukturen.

Markovsche Zufallsfelder Ein zufälliges Feld wird Markovsches Zufallfeld genannt, falls die bedingte Verteilung jedes einzelnen Pixels gegeben alle anderen Pixel nur von den Werten der Nachbarpixel abhängt,

$$\pi(x_i|\mathbf{x}_{-i}) = \frac{\pi(\mathbf{x})}{\sum_{x_i \in \mathcal{X}} \pi(\mathbf{x})} = \pi(x_i|\mathbf{x}_{\delta_i}) \ \forall \ i \in \mathcal{I}.$$

Bemerkungen

- Im diskreten Fall bedeutet die Markovfeldeigenschaft, dass

$$\begin{aligned}\mathbb{P}(X_i = x_i \mid \mathbf{X}_{-i} = \mathbf{x}_{-i}) &= \frac{\mathbf{P}(\mathbf{X}=\mathbf{x})}{\mathbf{P}(\mathbf{X}_{-i}=\mathbf{x}_{-i})} \\ &= \mathbb{P}(X_i = x_i \mid \mathbf{X}_{\delta i} = \mathbf{x}_{\delta i})\end{aligned}$$

- Markovsches Feld kann nur bedingt als Verallgemeinerung einer Markovkette angesehen werden, da keine eindeutige Richtung wie im linearen Fall.

Theorem von Hammersley-Clifford

Eine Verteilung $\pi = \{\pi(\mathbf{x}), \mathbf{x} \in \mathcal{X}^{|\mathcal{I}|}\}$, $\pi(\mathbf{x}) > 0$ für alle $\mathbf{x} \in \mathcal{X}^{|\mathcal{I}|}$ ist genau dann die Verteilung eines Markovschen Zufallsfeldes, falls sie folgende Form besitzt:

$$\pi(\mathbf{x}) = \frac{\exp\left(-\sum_{C\in\mathcal{C}} \phi_C(\mathbf{x}_C)\right)}{\sum_{\mathbf{x}} \exp\left(-\sum_{C\in\mathcal{C}} \phi_C(\mathbf{x}_C)\right)}$$

für gewisse Funktionen $\{\phi_C\}$, wobei $\mathcal{C}$ die Menge aller Cliques darstellt.

Beweis:
Eine Verteilung dieser Form erfüllt die Markoveigenschaft, da nur benachbarte Pixel betrachtet werden.
Gegenrichtung viel schwerer (Winkler,2003).

Das Hammersley-Clifford Theorem ermöglicht es anstatt die Verteilungen direkt anzugeben, sie über die Wahl der sogenannten *Potenzialfunktionen* $\{\phi_C\}$ zu definieren.

3 Bildmodelle: Ising- und Pottsmodell

Das Isingmodell für Binärbilder

- Binärbild, d.h. $x_i = 1$ (Schwarz) oder $x_i = 0$ (Weiss) (=> $\mathcal{X} = \{0, 1\}$).
- Vierernachbarschaft, Cliques sind Paare von benachbarten Pixeln.
- Idee: $\mathbb{P}(X_i \text{ schwarz} \mid k \text{ schwarze Nachbarn}) \propto \exp(\beta k)$.
- Potenzialfunktion:
  $\phi_C(\mathbf{x}_C) = -\beta * \mathbb{1}\{\text{Beide Pixel in C haben die gleiche Farbe}\},\ \beta > 0$.
- Gemeinsame Verteilung:
  $\pi(\mathbf{x}) = \frac{exp(\beta \sum_{i\sim j} I\{x_i = x_j\})}{\sum_{\mathbf{x}} exp(\beta \sum_{i\sim j} I\{x_i = x_j\})}$

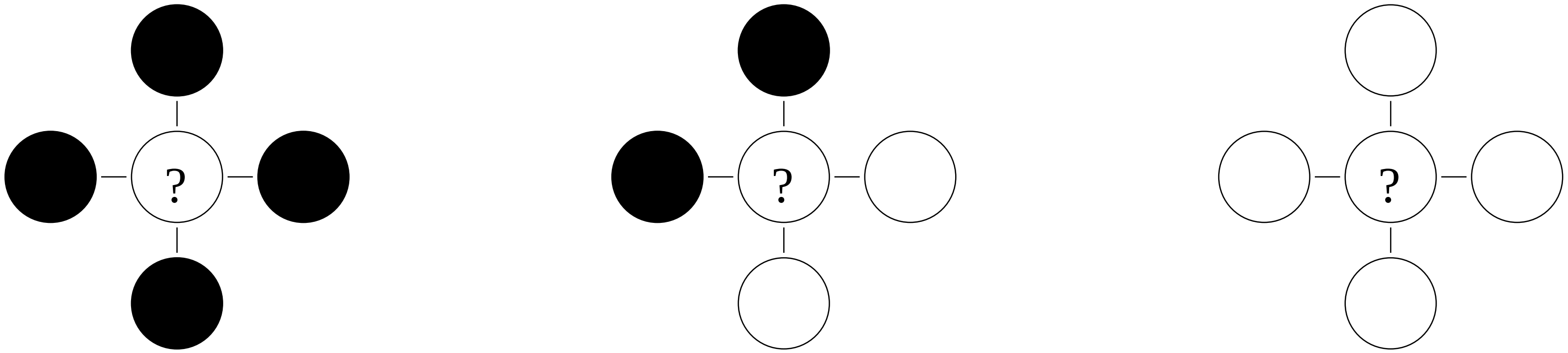

Abbildung 4: Verschiedene Nachbarschaftskonfigurationen.

Phasenübergang

Eine wichtige Eigenschaft des Isingmodells ist das sogenannte Phasenübergangsverhalten. Ausgehend von einem $n \times n$ Bild wird ein in der Mitte liegendes Pixel i^* betrachtet.

Es lässt sich nun zeigen, dass für $n \to \infty$ und $\beta > \beta_{crit} = \log(1 + \sqrt{2})$ die Verteilung von X_{i^*} durch die Werte am Rand des Bildes beeinflusst wird.

Daher ist es i.a. unklug, allein durch das Betrachten der bedingten Wahrscheinlichkeiten auf die globalen Modelleigenschaften zu schliessen.

Das Pottsmodell und andere Erweiterungen des Isingmodells

- Statt Pixel mit zwei möglichen Zuständen werden nun Pixel mit ℓ möglichen Zustände $\{0, 1, .., \ell - 1\}$ betrachtet.
- $\pi(\mathbf{x}) \propto \exp\left(\beta \sum_{i \sim j} \mathbb{1}\{x_i = x_j\}\right)$
- Gemeinsame Wahrscheinlichkeit hat dieselbe Gestalt, aber jetzt mehrere Zustände pro Einzelpixel anstatt binär.
- Bei Einteilung in gleichfarbige bzw. andersfarbige Nachbarn Rückkehr zum Isingmodell.
- Weitere Erweiterungsmöglichkeiten z.B. durch speziellere Wahl der β in Abhängigkeit des Abstands zwischen i und j und des angenommenen Zustands und/oder durch Wahl einer anderen Nachbarschaftsbeziehung (Achter, Zwölfer, ...).

4 MCMC Simulation von Bildmodellen

Gibbs Sampler für das Isingmodell

- Ausgangspunkt ist ein beliebig gewähltes Binärbild (z.B. $x_i = 1 \; \forall i \in \mathcal{I}$) mit $n \times n = N$ Pixeln.
- Seien $n_i^w = \sum_{j \sim i} \mathbb{1}\{x_j = 0\}$ und $n_i^b = \sum_{j \sim i} \mathbb{1}\{x_j = 1\}$.
- In jedem Iterationsschritt werden nacheinander die einzelnen Pixel x_i aufgerufen.
- Dieses Einzelpixel werden mit Wahrscheinlichkeit

$$p = \frac{\exp\left(\beta n_i^b\right)}{\exp\left(\beta n_i^b\right) + \exp\left(\beta n_i^w\right)}$$

auf 1 gesetzt, ansonsten auf 0.

Bemerkungen zum Gibbs Sampler

- Trotz seiner relativ simplen Struktur ist das Isingmodell π bereits hinreichend komplex um zur Simulation von π Markov-Chain-Monte-Carlo Methoden zu benötigen.
- Es wird eine Markovkette konstruiert mit π als stationärer Verteilung.
- Für das Aufrufen der einzelnen Pixel existieren verschiedene Möglichkeiten:
  - Numerische Ordnung : Im j-ten Schritt wähle Pixel j.
  - Zufällig: Wähle ein beliebiges Pixel.
  - Permutation. Wähle ein beliebiges Pixel aus denen, die noch nicht gewählt wurden.

Metropolis Sampler für das Isingmodell

- Gleicher Ausgangspunkt wie bei Gibbs Sampler.
- Wieder werden in jedem Iterationsschritt nacheinander die einzelnen Pixel x_i aufgerufen.
- Berechne

$$\begin{aligned} d &= \exp\left(\beta \textstyle\sum_{j\sim i} \mathbb{1}\{x_i = x_j\}\right) \\ d' &= \exp\left(\beta \textstyle\sum_{j\sim i} \mathbb{1}\{x_i \neq x_j\}\right). \end{aligned}$$

- Mit Wahrscheinlichkeit $p = \min\left\{1, \frac{d'}{d}\right\}$ wird das Pixel x_i verändert.

Bemerkungen zum Metropolis Sampler

- Erneut wird eine Markovkette mit stationärer Verteilung π konstruiert.
- Im Gegensatz zum Gibbs Sampler ist eine bessere Ausnutzung des Zustandsraums gewährleistet, da Wechsel von einer bestimmten Konfiguration zur inversen Konfiguration leichter möglich sind.
- Aus Gründen der numerischen Stabilität ist es ratsam einen etwas veränderten Ansatz zu wählen mit

$$\begin{aligned} d &= \beta \textstyle\sum_{j \sim i} \mathbb{1}\{x_i = x_j\} \\ d' &= \beta \textstyle\sum_{j \sim i} \mathbb{1}\{x_i \neq x_j\} \\ p &= \exp\left(\min\{0, d' - d\}\right. . \end{aligned}$$

Alternative Algorithmen

Falls $\beta > \beta_{crit}$ sind sowohl der Gibbs als auch der Metropolis Sampler aus Gründen der Konvergenzgeschwindigkeit nahezu ungeeignet.
Daher wurden alternative Algorithmen zur Simulation des Isingmodells entwickelt mit dem Algorithmus von Swendsen & Wang (1987) als bekanntestem Vertreter.
Hierbei wird eine hochdimensionale Variable $\mathbf{u}$ eingeführt, die für jede Nachbarschaftsbeziehung $i \sim j$ je eine Komponente u_{ij} enthält.
Nun kann unter gewissen Bedingungen eine Markovkette konstruiert werden, die abwechselnd Übergänge in $\mathbf{u}$ und $\mathbf{x}$ besitzt und $\pi(\mathbf{x})$ als stationäre Verteilung annimmt.

5 Gestörte Bilder und deren Simulation

- I.a. nicht möglich Bilder direkt zu beobachten, da sie gestört sind.
- Störungen können unvermeidlich sein oder durch Übertragungen entstehen.
- Statt des zugrunde liegenden Originalbilds $\mathbf{x}$ werden Daten $\mathbf{y}$ beobachtet.
- Im Folgenden werden Störungen für das Isingmodell betrachtet.

Bayesscher Ansatz bei gestörten Bildern

Bezeichnet man das Originalbild mit $\mathbf{x}$ und das beobachtete gestörte Bild mit $\mathbf{y}$ so erhält man unter Anwendung des Bayesschen Theorems

$$\pi(\mathbf{x} \mid \mathbf{y}) = \frac{\pi(\mathbf{y} \mid \mathbf{x})\pi(\mathbf{x}, \cdot)}{\pi(\cdot, \mathbf{y})} \propto \pi(\mathbf{y} \mid \mathbf{x})\pi(\mathbf{x}, \cdot)$$

mit $\pi : \mathcal{X}^{|\mathcal{I}|} \times \mathcal{Y}^{|\mathcal{J}|} \to \mathbb{R}_0^+$.

Bemerkungen

- In der Regel gilt $\mathcal{I} = \mathcal{J}$, aber $\mathcal{X}$ und $\mathcal{Y}$ können durchaus verschieden sein (z.B. $\mathcal{X} = \{0,1\}$ und $\mathcal{Y} = \mathbb{R}$).
- Die Randwahrscheinlichkeit $\pi(\cdot, \mathbf{y})$ ist für Aussagen über $\mathbf{x}$ nicht von Interesse.
- Die bedingte Wahrscheinlichkeit $\pi(\mathbf{y} \mid \mathbf{x})$ wird dazu verwendet um die Art der Störung zu modellieren.
- Eigentlich interessiert ist man an der a-posteriori Wahrscheinlichkeit $\pi(\mathbf{x} \mid \mathbf{y})$, die Aussagen über die Gestalt des Originalbilds ermöglicht.
- Im Folgenden entspricht $\pi(\mathbf{x}, \cdot)$ der Verteilung des Isingmodells.

Störungsarten

Seien Y_i und Y_j $(j \neq i)$ bedingt unabhängig für ein gegebenes $\mathbf{x}$.

Gausssche additive Störungen

$$\mathbf{Y} = \mathbf{X} + \varepsilon,$$

wobei ε normalverteilt ist mit Erwartungswertvektor $\mathbf{0}$, und Kovarianzmatrix $\sigma^2 I$. ε und $\mathbf{X}$ seien unabhängig.
Dann ist

$$\pi(y_i \mid x_i) \propto \frac{1}{\sigma} \exp\left(-\frac{(y_i - x_i)^2}{2\sigma^2}\right)$$

und

$$\pi(\mathbf{y} \mid \mathbf{x}) = \prod_{i \in \mathcal{I}} \pi(y_i \mid x_i).$$

Für die a-posteriori Wahrscheinlichkeit ergibt sich somit

$$\pi(\mathbf{x} \mid \mathbf{y}) \propto \exp\big(\beta \sum_{i \sim j} \mathbb{1}\{x_i = x_j\} + \sum_i h_i(x_i, y_i)\big)$$

mit

$$h_i(x_i, y_i) = -\frac{(y_i - x_i)^2}{2\sigma^2}.$$

Allgemein lässt sich jede Art von bedingt unabhängiger Störung in obiger Form schreiben, wobei

$$h_i(x_i, y_i) = \log(\pi(y_i \mid x_i))$$

ist.

Binäre Störungen

Jedes Einzelpixel wechselt mit Wahrscheinlichkeit p die Farbe.

$$\pi(y_i \mid x_i) = \begin{cases} 1-p & \text{falls} \quad y_i = x_i \\ p & \text{falls} \quad y_i \neq x_i \end{cases}$$

Es gilt erneut

$$\pi(\mathbf{x} \mid \mathbf{y}) \propto \exp\big(\beta \sum_{i \sim j} \mathbb{1}\{x_i = x_j\} + \sum_i h_i(x_i, y_i)\big),$$

dieses Mal mit

$$h_i(x_i, y_i) = \mathbb{1}\{y_i = x_i\} \log\big(\frac{1-p}{p}\big).$$

Weitere Bemerkungen zu den Störungsarten

- Es existieren noch viele andere Arten von Störungen.
- Wichtig ist,dass die y_i bei additiven Störungen nicht mehr binär sein müssen, sondern unter Umständen kontinuierlich abgebildet werden.
- Bei binären Störungen enthält für $p = 1/2$ das beobachtete Bild keine Information mehr über das Originalbild.

Simulation ausgehend von der a-posteriori Verteilung

- A-posteriori Wahrscheinlichkeit enthält Informationen über das Originalbild basierend auf den beobachteten Daten und dem a-priori Wissen.
- Daher können nach den a-posteriori Wahrscheinlichkeiten verteilte Stichproben Aussagen über das Originalbild ermöglichen.
- Algorithmen für ungestörte Bildmodelle müssen nur leicht modifiziert werden um den Einfluss der a-priori Annahmen zu berücksichtigen.
- Bessere Konvergenzgeschwindigkeiten, da Problem des Phasenübergangs durch Information über die beobachteten Daten eliminiert wird.

Metropolis Sampler für das gestörte Isingmodell

- Initialisere **x**, gegeben **y** und Störungsparameter.
- Es werden in jedem Iterationsschritt nacheinander die einzelnen Pixel x_i aufgerufen.
- Berechne
  $d = \beta \sum_{j \sim i} \mathbb{1}\{x_i = x_j\} + h_i(x_i, y_i)$,
  $d' = \beta \sum_{j \sim i} \mathbb{1}\{x_i \neq x_j\} + h_i(x_i, y_i)$.
- Mit Wahrscheinlichkeit $p = \exp(\min\{0, d' - d\})$ wird das Pixel x_i verändert.

6 Schätzung von Bildern

- Meist interessiert man sich nicht nur für die a-posteriori Verteilung eines beobachteten Bildes, sondern möchte daraus einen optimalen Schätzer für das Originalbild $\mathbf{x}^*$ ermitteln, das zu dem beobachteten Bild $\mathbf{y}^*$ gehört.
- Hierzu ist es nötig ein Abstandsmaß $L(\mathbf{x}, \mathbf{z})$ zwischen zwei Bildern $\mathbf{x}$ und $\mathbf{z}$ zu definieren.
- Dieses Abstandsmaß muss nicht notwendigerweise im mathematischen Sinn eine Metrik sein.
- Wichtiger: Gute Wiedergabe des (subjektiven) Bildabstands.

- Der mitttlere a-posteriori Abstand zwischen einer Schätzung $\mathbf{z}$ und dem unbekannten $\mathbf{x}$ wird definiert als

$$E_{\mathbf{x}|\mathbf{y}} L(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{x}} \pi(\mathbf{x} \mid \mathbf{y}) L(\mathbf{x}, \mathbf{z})$$

- Der optimale Bayessche Schätzer für $\mathbf{x}$ ist dann gegeben durch

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{z}} E_{\mathbf{x}|\mathbf{y}} L(\mathbf{x}, \mathbf{z})$$

Beispiele für Bayessche Schätzer: (bzgl. gestörtem Isingmodell)

Marginal-Posterior-Mode (MPM) Schätzer

Abstandsmaß

$$L_{MPM}(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{N} \mathbb{1}\{x_i \neq z_i\}$$

daher

$$\mathbf{x}_{MPM} = \arg\min_{\mathbf{z}} \left\{ -\sum_i \mathbb{P}(X_i = z_i \mid \mathbf{y}) \right\} = \sum_i \arg\max_{z_i} \mathbb{P}(X_i = z_i \mid \mathbf{y})$$

und

$$\mathbf{x}_{MPM,i} = \begin{cases} 1 & \text{falls} \quad \mathbb{P}(X_i = 1 \mid \mathbf{y}) > 1/2 \\ 0 & \text{falls} \quad \mathbb{P}(X_i = 1 \mid \mathbf{y}) \leq 1/2 \end{cases}$$

Bemerkungen zum MPM Schäetzer

- Der MPM Schäetzer minimiert die Zahl der Fehlklassifikationen, allerdings ohne Berücksichtigung der Lage.
- Zur Berechnung von $\mathbf{x}_{MPM}$ werden K Realisierungen der a-posteriori Verteilung erzeugt.Dann ergibt sich für $\mathbf{x}_{MPM,i}$:

$$\mathbf{x}_{MPM,i} = \begin{cases} 1 & \text{falls} \quad \#\{\text{simulierte } x_i \text{ mit } x_i = 1\} > K/2 \\ 0 & \text{falls} \quad \#\{\text{simulierte } x_i \text{ mit } x_i = 1\} \leq K/2 \end{cases}$$

Maximum-A-Posteriori (MAP) Schätzer

Abstandsmaß

$$L_{MAP}(\mathbf{x},\mathbf{z}) = 1 - \prod_{i=1}^{N} (1 - \mathbb{1}\{x_i \neq z_i\})$$

daher

$$\mathbf{x}_{MAP} = \arg\max_{\mathbf{z}} \pi(\mathbf{z} \mid \mathbf{y})$$

Bemerkungen zum MAP Schätzer

- MAP Schätzer unterscheidet nicht zwischen verschiedenen Wertigkeiten bei falschen Bildern.
- Schätzer entspricht Modus der a-posteriori Wahrscheinlichkeit.
- Einfache Schätzmethode ist Simulation und Speicherung der a-posteriori Wahrscheinlichkeit. Allerdings sehr ineffizient.
- Bessere Schätzmethode mittels MCMC und Simulated Annealing.

7 Literatur

- J. Møller. *Spatial Statistics and Computational Methods*, Lecture Notes in Statistics 173, Springer, 2003.
- B. Chalmond. *Modeling and Inverse Problems in Image Analysis*, Springer-Verlag, 2003.
- G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, Springer-Verlag, 2003.
- R. Swendsen & J. Wang. *Nonuniversal Critical Dynamics in Monte Carlo Simulations*, Physical Review Letters **58**: 86-88, 1997.