



Seminar Information Retrieval

Tipps zu Ausarbeitung und Vortrag

Steffen Bickel

Agenda

- Schriftliche Ausarbeitung
- Vortragspräsentation

Agenda

- Schriftliche Ausarbeitung
- Vortragspräsentation

Literatursuche

- Ausgangsliteratur vorgegeben
- Eigene Literaturrecherche
 - ◆ Interessante Verweise aus „Related Work“ verfolgen
 - ◆ Neue Veröffentlichungen, die evtl. auf Ausgangsliteratur verweisen.
- Literatursuche im Internet:
 - ◆ <http://www.scholar.google.com>
 - ◆ <http://citeseerx.ist.psu.edu>

Literatursuche mit <http://scholar.google.com>

[Advanced Scholar Search](#)[Scholar Preferences](#)[Scholar Help](#)

Scholar All articles - [Recent articles](#)

[Multi-view clustering](#) - [all 12 versions](#) »

S Bickel, T Scheffer - *Proceedings of the IEEE International Conference on Data ...*, 2004 - 141.20.20.55

Page 1. **Multi-View Clustering** ... We develop and study partition- ing and agglomerative, hierarchical **multi-view clustering** algorithms for text data. ...

[Cited by 18](#) - [Related Articles](#) - [View as HTML](#) - [Web Search](#) - [Import into BibTeX](#)

[Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?"](#) - [all 7 versions](#) »

F Schaffalitzky, A Zisserman - *Proc. ECCV, 2002* - Springer

... **Multi-view** Matching for Unordered Image Sets 429 Fig. 12. The largest **cluster** (8 views), for the Raglan image set, showing the interior of the hexagonal keep. ...

[Cited by 158](#) - [Related Articles](#) - [Web Search](#) - [Import into BibTeX](#) - [BL Direct](#)

[3D object modeling and recognition using affine-invariant patches and **multi-view** spatial constraints](#) - [all 13 versions](#) »

F Rothganger, S Lazebnik, C Schmid, J Ponce - *Computer Vision and Pattern Recognition, 2003. Proceedings. ...*, 2003 - ieeexplore.ieee.org

Page 1. 3D Object Modeling and Recognition Using Affine-Invariant Patches and **Multi-View** Spatial Constraints Fredrick Rothganger ...

[Cited by 56](#) - [Related Articles](#) - [Web Search](#) - [Import into BibTeX](#) - [BL Direct](#)

[M 2 Tracker: A **Multi-View** Approach to Segmenting and Tracking People in a Cluttered Scene](#) - [all 16 versions](#) »

A Mittal, LS Davis - *International Journal of Computer Vision*, 2003 - Springer

... Page 3. M 2 Tracker: A **Multi-View** Approach to Segmenting and Tracking People in a Cluttered Scene ... **Clustering** these points allows us to detect and track people. ...

[Cited by 121](#) - [Related Articles](#) - [Web Search](#) - [Import into BibTeX](#) - [BL Direct](#)

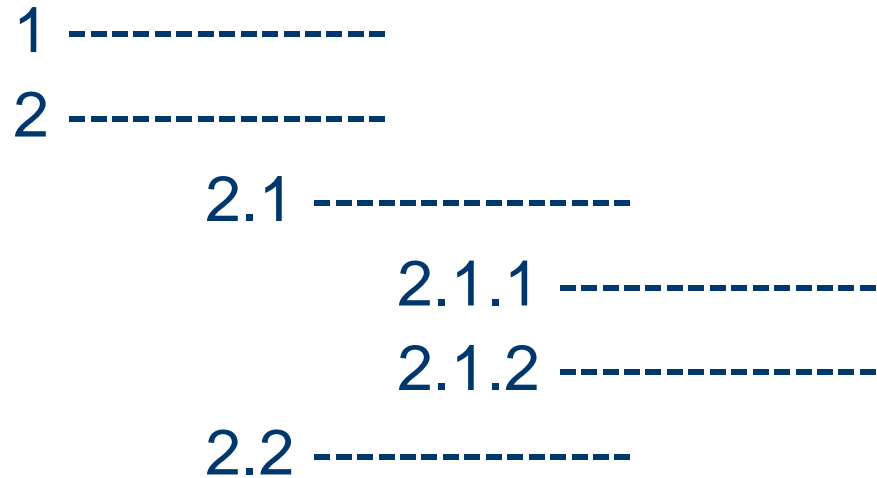
[Active+ semi-supervised learning= robust **multi-view** learning](#) - [all 10 versions](#) »

I Muslea, S Minton, C Knoblock - *Proceedings of the International Conference on Machine ...*, 2002 - aicml.cs.ualberta.ca

... **clustering** the unlabeled data "around" the ex- amples in the original training set.

3.3 Co-EM Co-EM(Nigam & Ghani, 2000) is a semi-supervised, **multi-view** ...

Gliederung – Gliederungstiefe



- Gliederungstiefe maximal drei Ebenen

Gliederung - Abstraktionsniveau

1 Verkehrsmittel

- 1.1 Auto
- 1.2 Flugzeug
- 1.3 Bahn
- 1.4 ICE
- 1.5 Nahverkehr

besser

1 Verkehrsmittel

- 1.1 Auto
- 1.2 Flugzeug
- 1.3 Bahn
 - 1.3.1 ICE
 - 1.3.2 Nahverkehr

2 Nutzungsvorraussetzung

- 2.1 Schaffner
- 2.2 Pilot
- 2.3 Straßennetz
- 2.4 Flughafen
- 2.5 Schienennetz

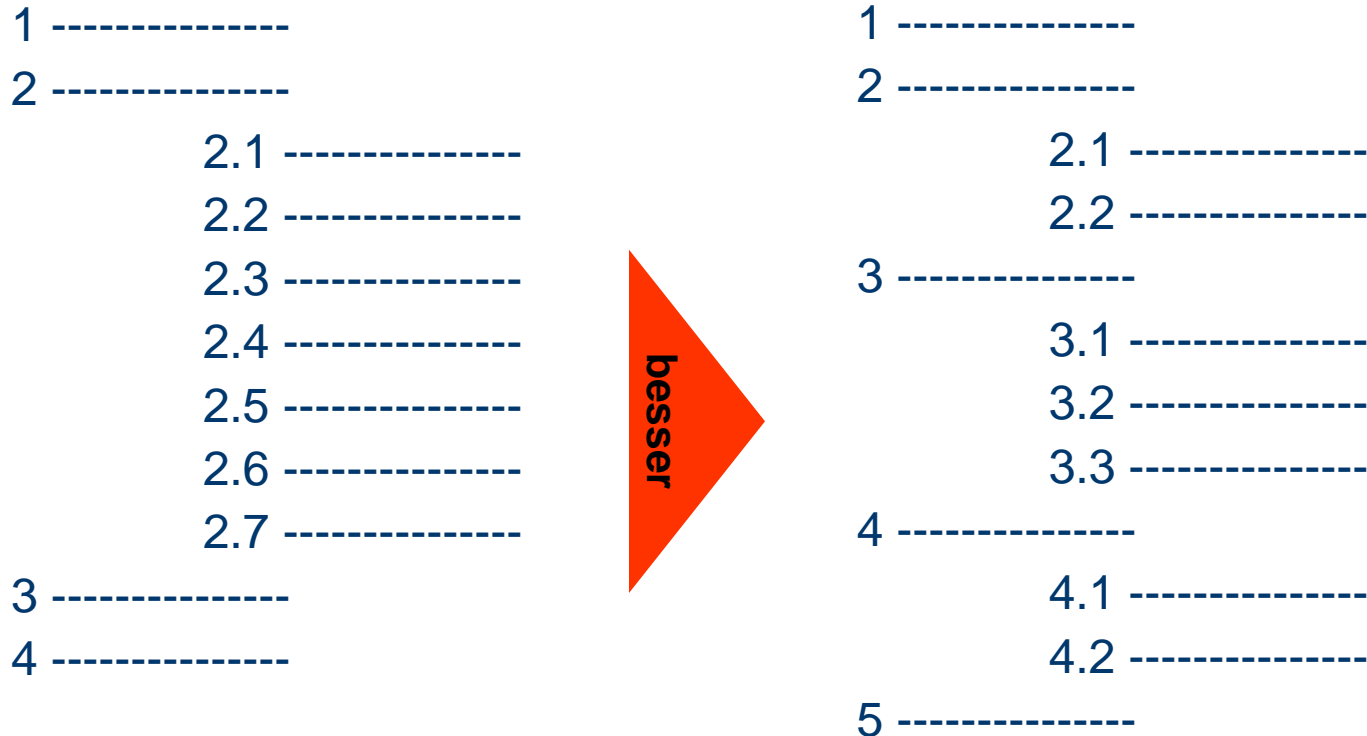
besser

2 Nutzungsvorraussetzung

- 2.1 Personal
 - 2.1.1 Schaffner
 - 2.1.2 Pilot
- 2.2 Infrastruktur
 - 2.2.1 Straßennetz
 - 2.2.2 Flughafen
 - 2.2.3 Schienennetz

- Abschnitte auf derselben Gliederungsebene sollten sich inhaltlich auf dem gleichen Abstraktionsniveau befinden

Gliederung – Strukturierung



- Auf jeder Ebene nicht mehr als sechs Abschnitte hintereinander, optimal: 2-4 Abschnitte

Gliederung – Keine Einzelnen Kapitel

1 -----
2 -----
Text A
2.1 -----
Text B
3 -----



1 -----
2 -----
Text A
Text B
3 -----



1 -----
2 -----
2.1 -----
Text A
2.2 -----
Text B
3 -----

Gliederung – Keine Kapitelüberschrift Identisch mit Ausarbeitungstitel

Beispiel-Titel: Vorhersage von Fussballergebnissen mit Probabilistischen Modellen

1 Einleitung

2 Fussball

2.1 -----

2.2 -----



3 Vorhersage von Fussballergebnissen mit Probabilistischen Modellen

3.1 -----


3.2 -----

4 Zusammenfassung

Gliederungsüberschrift – Verben und Artikel

- Verben substantivieren
- Kein Artikel als 1.Wort
- Beispiel:
 - ◆ „Die Überanpassung vermeiden“
 - ◆ besser: „Vermeidung von Überanpassung“

Ausarbeitung – Allgemeine Gliederung

- Titelblatt
 - Inhaltsverzeichnis
 - evtl. Abbildungs-/Tabellen-/
Symbol-/Abkürzungsverzeichnis
 - Einleitung
 - Hauptteile
 - Zusammenfassung/Schlussfolgerung
 - Literaturverzeichnis
 - evtl. Anhang
- 
- in Inhaltsverzeichnis

Ausarbeitung – Tabellen, Diagramme, Formeln

- Tabellen und Abbildungen immer mit fortlaufender Nummerierung und Bezeichnung
- z.B. *Abbildung 5: Fehlerrate des Textfilter-Algorithmus*
- Formeln mit rechts- oder linksbündiger Nummerierung z.B.

$$Entropy(S) = \sum_{i=1}^c - p_i \log_2 p_i \quad (2)$$

- Im fortlaufenden Text auf Abbildungen oder Formeln verweisen

Ausarbeitung - Stil

- Keine Umgangssprache
- Keine rhetorischen Fragen
- Keine subjektiven Bewertungen

Wie lernt man den Stil wissenschaftlicher Arbeiten?

- ➔ Andere wissenschaftliche Arbeiten lesen und den Stil und „Klang“ übernehmen

Ausarbeitung – Englische Fachbegriffe

- Feststehende Fachbegriffe nicht übersetzen (z.B. Data Mining)
- Falls vorhanden, lieber deutsches Äquivalent verwenden (z. B. „Überanpassung“ statt „Overfitting“)
- Falls keine adäquate Übersetzung vorhanden, keine gewagten Wortschöpfungen (nicht „Stützvektormaschine“ für „Support Vector Machine“)

Ausarbeitung – Zitate und Quellen

- Jede übernommene Idee und jeder Textauszug müssen als Zitate erkennbar sein
- Kein Zitat ohne Quelle
- Nicht gekennzeichnete Zitate sind **Plagiate**
- In der Informatik hat sich die amerikanische Zitierweise durchgesetzt:
 - ◆ „... mit QP-Methode gelöst werden [13].“
 - ◆ „... mit QP-Methode gelöst werden (Mitchell, 2001).“
 - ◆ „Mitchell (2001) löst dies mit der QP-Methode.“

Ausarbeitung - Literaturverzeichnis

- Das Literaturverzeichnis sortieren nach:
 - ◆ Autornamen oder
 - ◆ Auftretensreihenfolge im Text.
- Mehrere Veröffentlichungen eines Autors in einem Jahr:
 - ◆ Brown 2002a, Brown 2002b, ...
- Bibtex für Latex macht dies automatisch.

Ausarbeitung - Allgemeine Hinweise

- Abgabe der Ausarbeitung als MS-Word- oder PDF-Datei per Email an jeweiligen Betreuer.
- Ausarbeitung ca. 12-15 Seiten (Inhalt zählt, nicht Anzahl der Seiten).
- DIN A4, deutsch oder englisch.
- Dokumentvorlagen für Latex gibt es im Internet im Überfluss.
- Vorteile Latex:
 - ◆ plattformunabhängig,
 - ◆ wissenschaftlicher Standard,
 - ◆ professionelles Schriftbild, besonders für mathematische Ausdrücke,
 - ◆ Bibtex für Quellen-Management.

Agenda

- Schriftliche Ausarbeitung
- Vortragspräsentation

Vortrag - Allgemeine Hinweise

- Vortragsdauer
 - ◆ 20 Min.
- Zeitlimit einhalten, nicht zu kurz, nicht zu lang.
- 1 Folie \approx 1-1,5 Minuten, ergibt 14-20 Folien
- Vortrag vorher gut üben, Zeit messen.
- **Wichtig:** Zielgruppe sind Mitstudenten – nicht Betreuer.

Aufbau – Grundlegendes

- Einleitung
- Hauptteil (Was will ich eigentlich sagen?)
- Zusammenfassung

Aufbau – Einleitung

- Titelfolie auflegen, Begrüßung, Vorstellung,
- Definition & Motivation des Problems
 - ◆ Interesse wecken!
- Evtl. Beziehung zu vorhergehenden Vorträgen herstellen.
- Agenda präsentieren.

Hauptteil

- Logische Gliederung, Roter Faden muss sich klar durchziehen.
- Wichtigste Fakten, keine unwichtigen Details.
- Wichtige Aussagen wiederholen.

Zusammenfassung und Ausblick

- Inhalt kurz & prägnant zusammenfassen
 - ◆ Letzte Folie: “Take Home Message” von ca. 3 Punkten
- Evtl. ungelöste Probleme ansprechen.
- Evtl. fortführende Arbeiten nennen.

- Ausgangspunkt für die Diskussion
- Zuhörer ermuntern, Fragen zu stellen

Foliensstil

- Große, lesbare Schrift (mind. 16pt)
- Stichworte besser als ganze Sätze
- Keine schwer nachvollziehbaren Beweise
- Übersichtlichkeit
 - ◆ Wenige Aufzählungspunkte pro Folie (max. 6)
 - ◆ Keine umfangreichen Quellcodeauszüge
 - ◆ Keine komplexen Tabellen, Diagramme
- Farbe & Animationen gezielt einsetzen.
- Ein Bild sagt mehr als 1000 Worte.

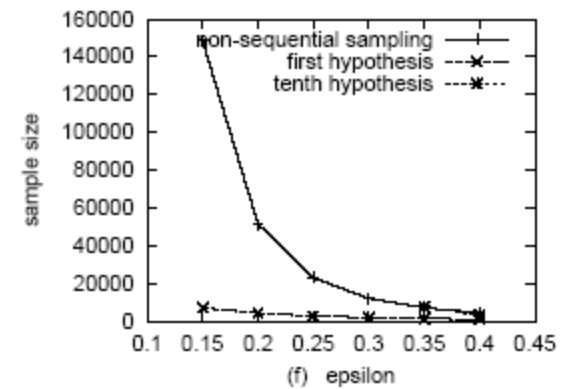
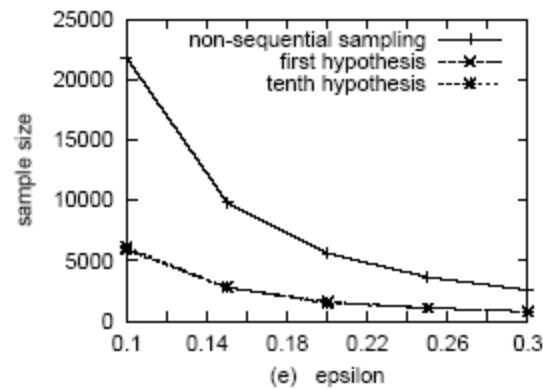
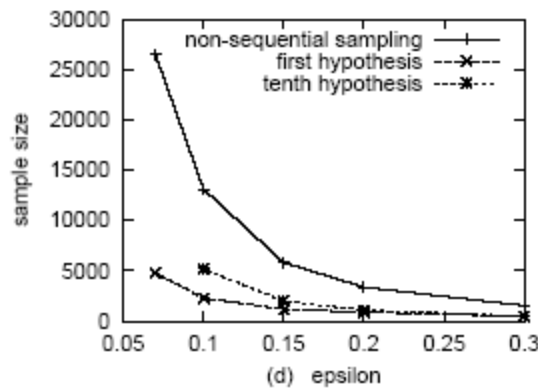
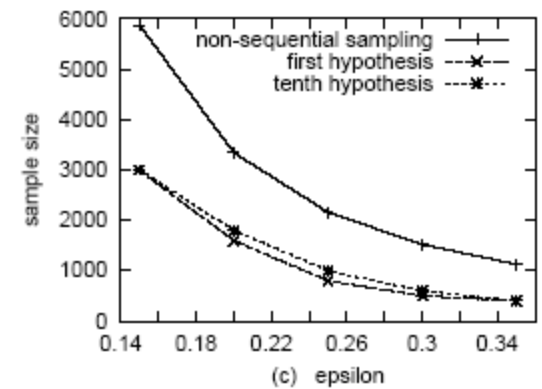
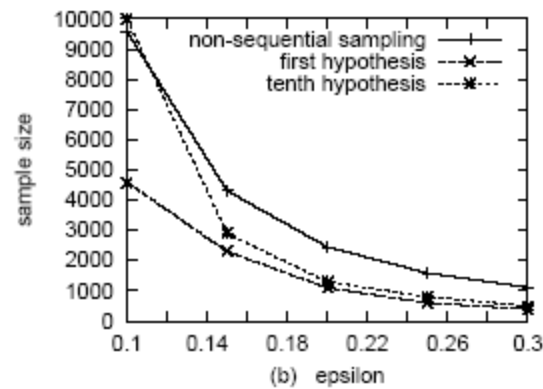
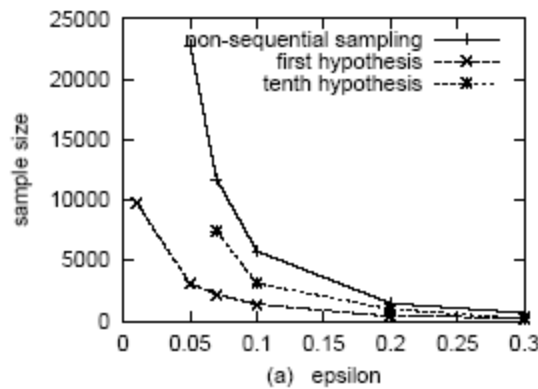
Schriftgröße mind. 16 Punkte, je größer desto besser

- Das Auto frisst die Katze (28 pt)
- Das Auto frisst die Katze (24 pt)
- Das Auto frisst die Katze (20 pt)
- Das Auto frisst die Katze (16 pt)
- Das Auto frisst die Katze (12 pt)
- Das Auto frisst die Katze (10 pt)

Falsch: Zu unübersichtlich

Year	Winning Candidate	Gallup Final Survey	Election Result	Deviation
1936	Roosevelt	55.7%	62.5%	6.8%
1940	Roosevelt	52.0%	55.0%	3.0%
1944	Roosevelt	51.5%	53.3%	1.8%
1948	Truman	44.5%	49.9%	5.4%
1952	Eisenhower	51.0%	55.4%	4.4%
1956	Eisenhower	59.5%	57.8%	1.7%
1960	Kennedy	51.0%	50.1%	0.9%
1964	Johnson	64.0%	61.3%	2.7%
1968	Nixon	43.0%	43.5%	0.5%
1972	Nixon	62.0%	61.8%	0.2%
1976	Carter	48.0%	50.0%	2.0%
1980	Reagan	47.0%	50.8%	3.8%
1984	Reagan	59.0%	59.1%	0.1%
1988	Bush	56.0%	53.9%	2.1%
1992	Clinton	49.0%	43.2%	5.8%
1996	Clinton	52.0%	50.1%	1.9%

Falsch: Zu unübersichtlich



Sample sizes for the juice purchases database. (a) $f = g|p - p_0|$, $k = 1$, $\delta = .1$; (b) $k = 2$; (c) $k = 3$; (d) $f = g^2|p - p_0|$, $k = 1$, $\delta = .1$; (e) $k = 2$; (f) $f = \sqrt{g}|p - p_0|$, $k = 1$, $\delta = .1$

Falsch: Zu unübersichtlich

Theorem 8 *Let*

1. $f(h) = g^2(p - p_0)$ and $\hat{f}(h, Q) = \hat{g}^2(\hat{p} - p_0)$ or
2. $f(h) = g^2|p - p_0|$ and $\hat{f}(h, Q) = \hat{g}^2|\hat{p} - p_0|$ or
3. $f(h) = g^2 \frac{1}{c} \sum_{i=1}^c |p_i - p_{0i}|$ and $\hat{f}(h, Q) = \hat{g}^2 \frac{1}{c} \sum_{i=1}^c |\hat{p}_i - p_{0i}|$ or

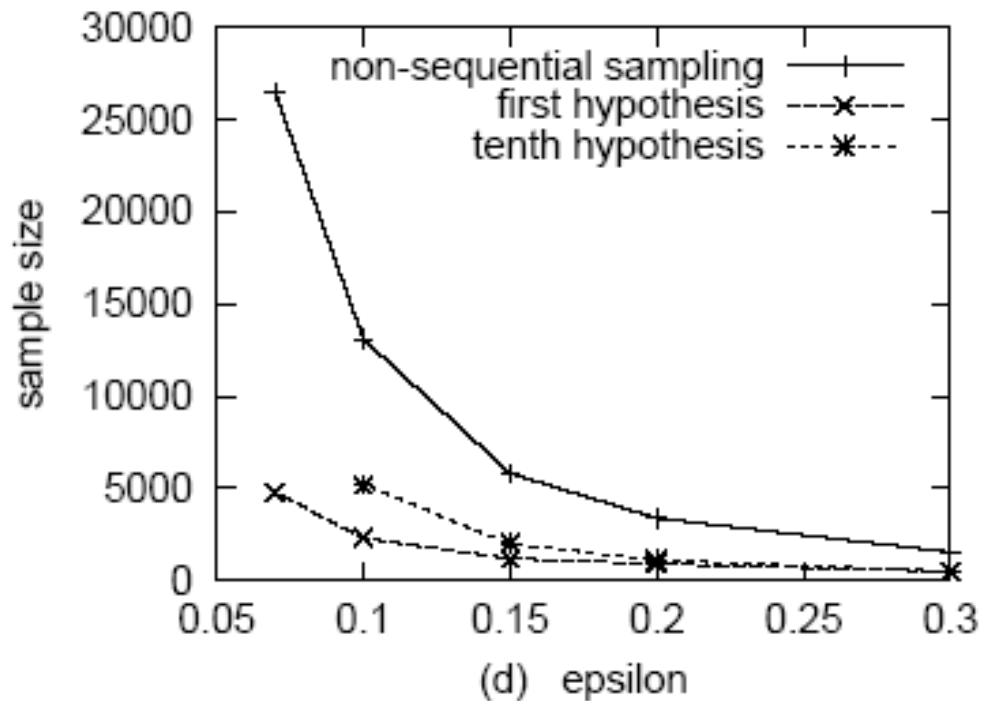
Then $\Pr[\hat{f}(h, Q_m) - f(h, Q) \leq E(m, \delta)] \geq 1 - \delta$ when

$$\text{small } m : \quad E(m, \delta) = \left(\frac{1}{2m} \log \frac{4}{\delta} \right)^{\frac{3}{2}} + 3 \left(\frac{1}{2m} \log \frac{4}{\delta} \right) + 3 \sqrt{\frac{1}{2m} \log \frac{4}{\delta}} \quad (33)$$

$$\text{large } m : \quad E(m, \delta) = \frac{3}{2\sqrt{m}} z_{1-\frac{\delta}{2}} + \frac{m + \sqrt{m}}{4m\sqrt{m}} (z_{1-\frac{\delta}{2}})^2 + \frac{1}{8m\sqrt{m}} (z_{1-\frac{\delta}{2}})^3 \quad (34)$$

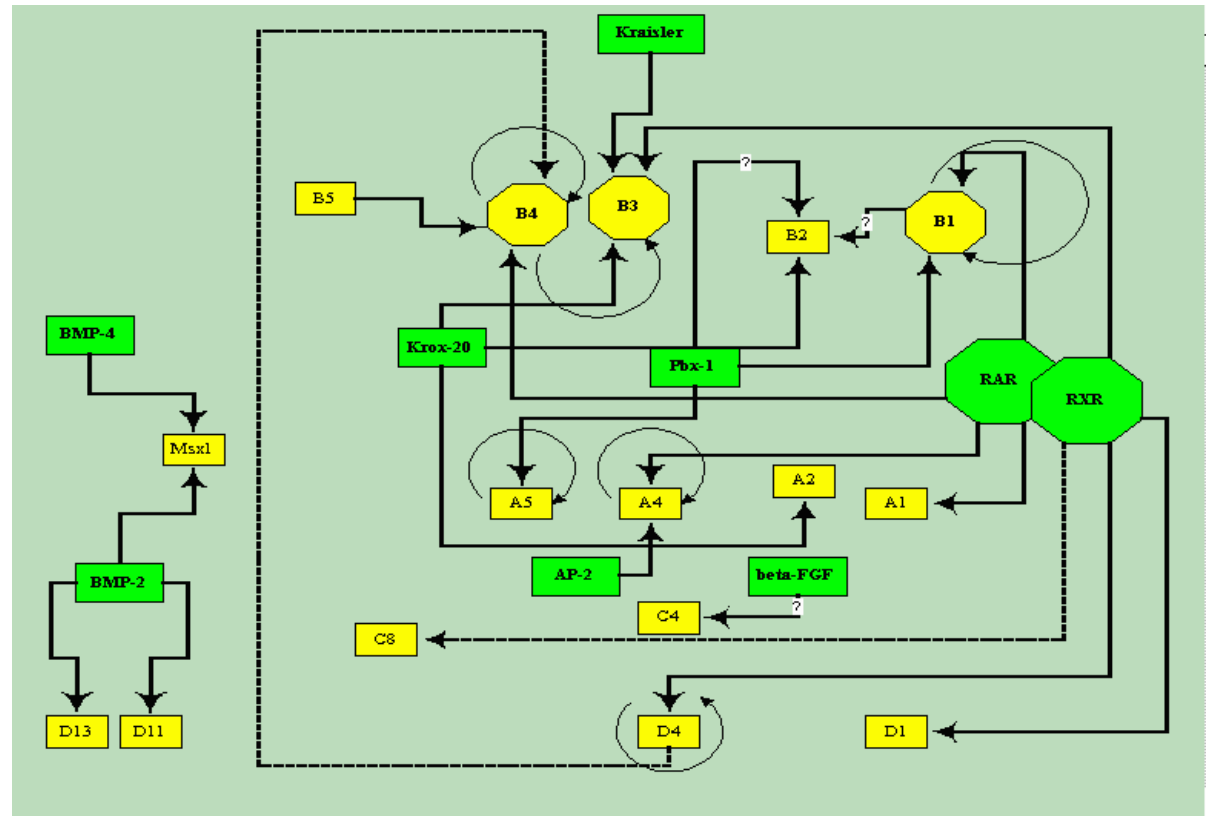
$$E_h(m, \delta) = 2s_g z_{1-\frac{\delta}{2}} + s_g^2 (z_{1-\frac{\delta}{2}})^2 + s_p z_{1-\frac{\delta}{2}} + 2s_g s_p (z_{1-\frac{\delta}{2}})^2 + s_p s_g^2 (z_{1-\frac{\delta}{2}})^3 \quad (35)$$

Richtig: Ergebnis im Vordergrund



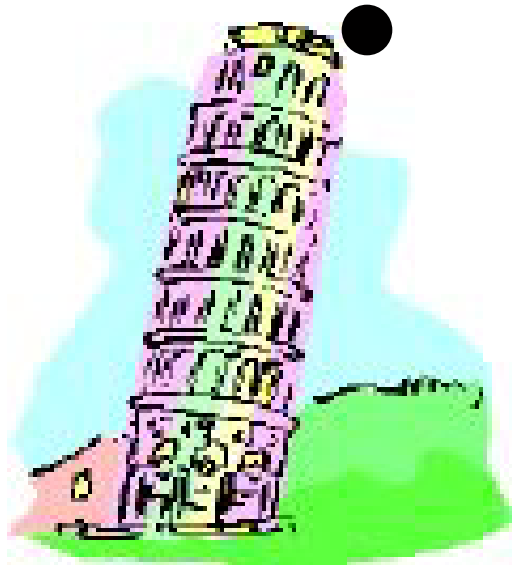
- Sequentielles Sampling um Faktor 2-10 schneller als nicht-sequentielles Sampling.

Bilder und Animationen: Falsch



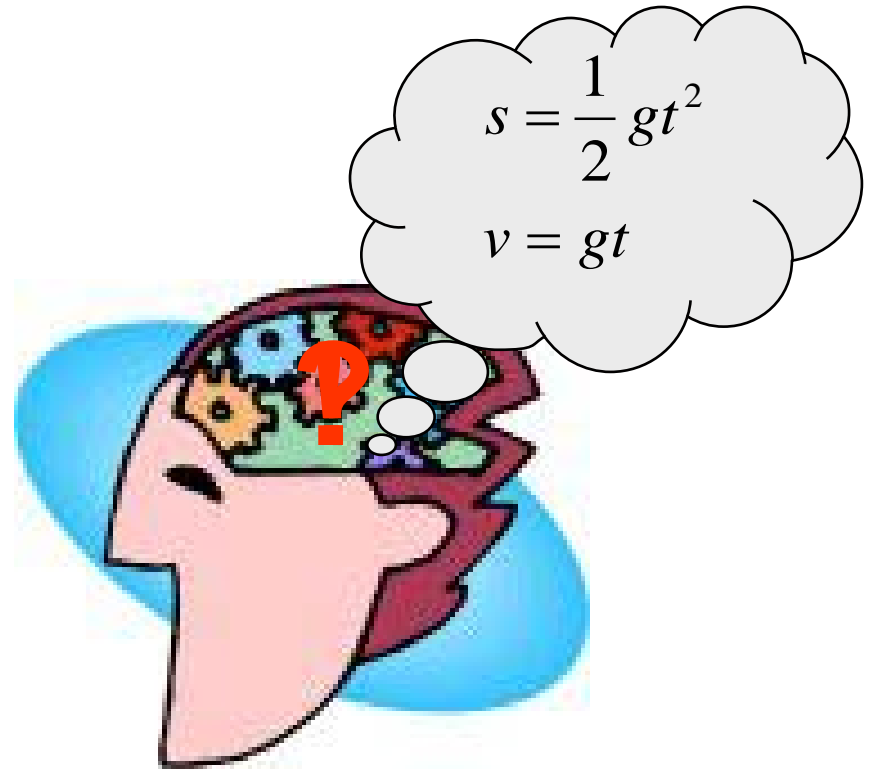
- Keine nutzlosen Animationen
- Grafik & Animation dienen dem Inhalt, nicht umgekehrt.

Bilder und Animationen: Richtig Maschinelles Lernen



Umgebung

Daten



Lern-Algorithmus

Modell

Vortragstil

- Möglichst frei sprechen, kein Ablesen
 - ◆ Blickkontakt mit Zuhörern
 - ◆ Auf Verständnisprobleme achten
 - ◆ Nicht auf Projektionsfläche sehen oder zeigen
- Keine Entschuldigungen, keine Füllwörter “Ääh”.
- Keine tief verschachtelten oder halben Sätze.
- Ruhig stehen, nicht rumzappeln.
- Ruhig gestikulieren
 - ◆ Zettel, Stift oder Maus hält Hände davon ab, Eigenleben zu entwickeln
- Humor lockert Stimmung auf.

Präsentation – Technisches

- Folien mit Powerpoint oder Open Office
- Sicherer Fallback
 - ◆ Overhead-Folien/Folien auf Memorystick
 - ◆ Stichwortzettel aus Papier
- Technik:
 - ◆ Eigener Laptop oder
 - ◆ Laptop von uns
 - ★ Acrobat Reader
 - ★ Power Point
 - ★ Folien auf Memory Stick mitbringen oder per Email an uns.
 - ◆ Beamer vorhanden

Take Home Message

- Zielgruppe sind Mitstudenten.
- Vortrag nicht notwendiges Übel, sondern Chance unter wohlgesinnten Zuhörern Vortragstechnik zu üben und verbessern.
- Zeit gut einteilen, nicht alles auf den letzten Drücker.

- Wir wünschen euch viel Erfolg!