Dirichlet Process Mixture Models for Object Category Recognition

Sandra Ebert

ICVSS 2009

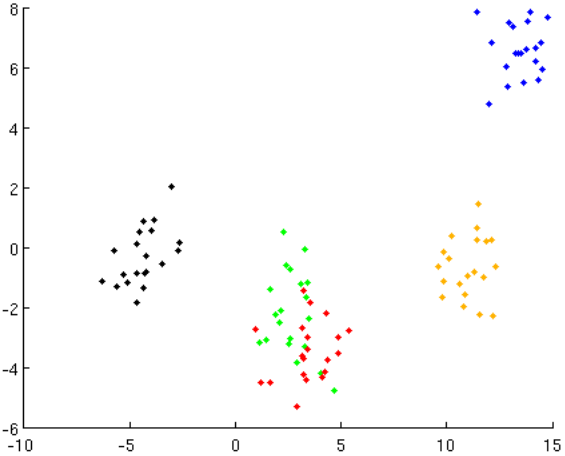
08. Jul 2009

Adviser: Michael Stark, Mario Fritz, Bernt Schiele

Outline

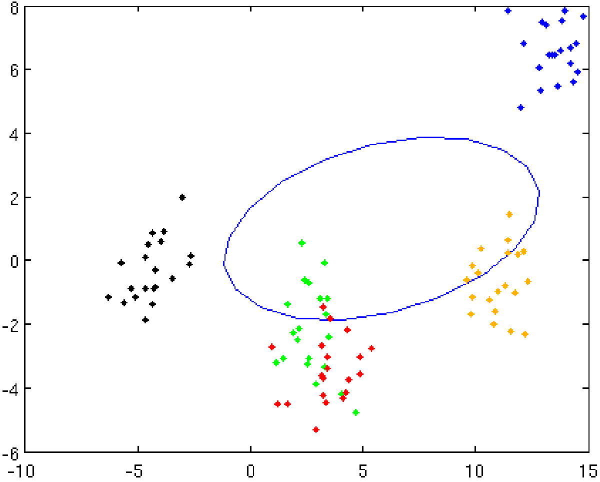
Challenge: Model Selection

Find the best model for the following training set to classify new and unseen data: How many mixtures (cluster)?



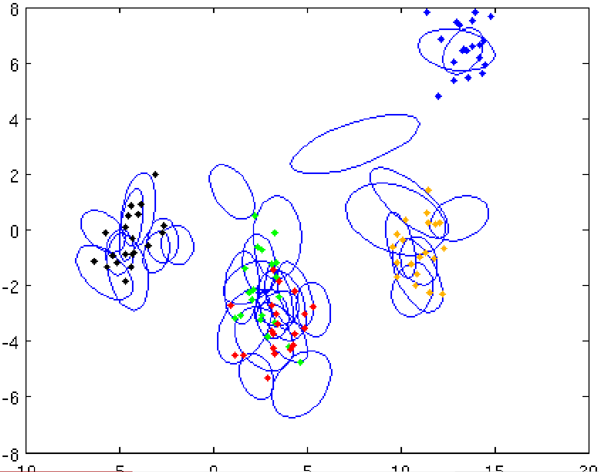
Model Selection: Under-Fitting

Only few clusters ⇒ the model is too general and the accuracy is poor

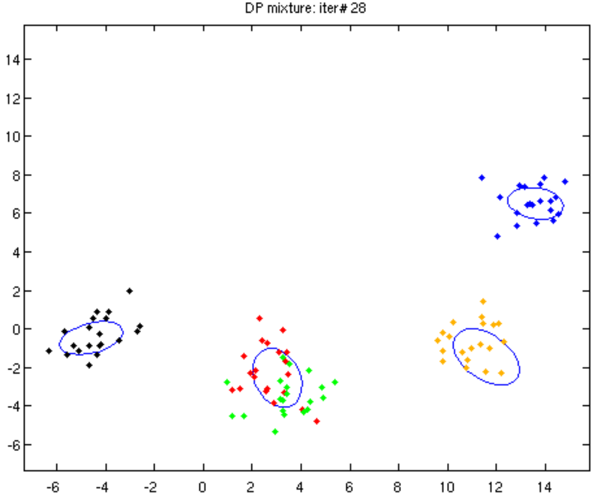


Model Selection: Over-Fitting

Nearly one cluster for each data point ⇒ the model is too specific and fails for new and unseen data

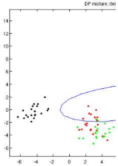


Model Selection: Good solution

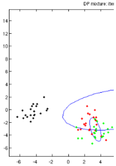


Model Selection with parametric approaches

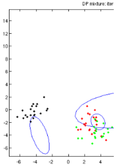
Try several models with different parameters and choose the best model according to a measure such as accuracy or log-likelihood



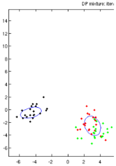
k=1
acc=20%
loglik = -603.1



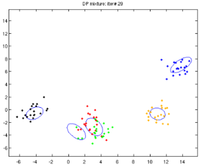
k=2
acc=60.2%
loglik = -600.8



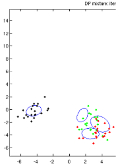
k=3
acc=60.5%
loglik = -593.8



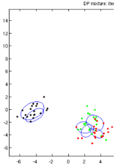
k=4
acc=87.5%
loglik = -329.1



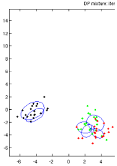
k=5
acc=90.3%
loglik = -330.5



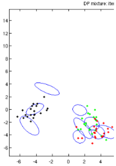
k=6
acc=85.7%
loglik = -329.9



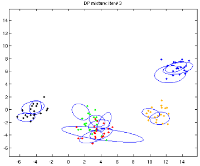
k=7
acc=80.7%
loglik = -318.3



k=9
acc=84.4%
loglik = -333.9



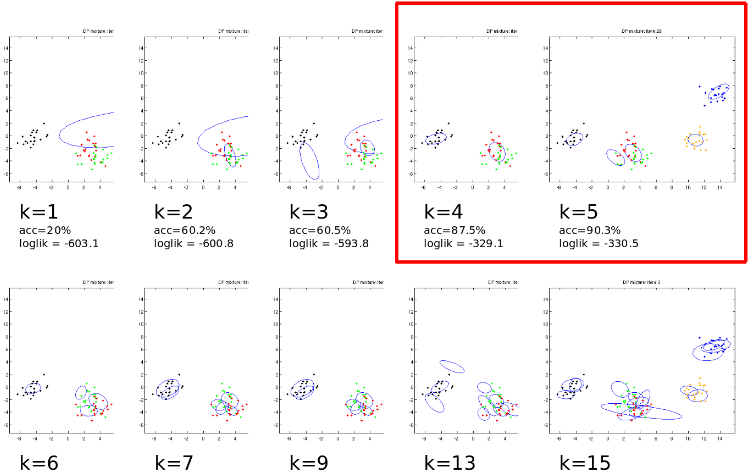
k=13
acc=84.9%
loglik = -342.9



k=15
acc=80.7%
loglik = -364.1

Model Selection with parametric approaches

Try several models with different parameters and choose the best model according to a measure such as accuracy or log-likelihood

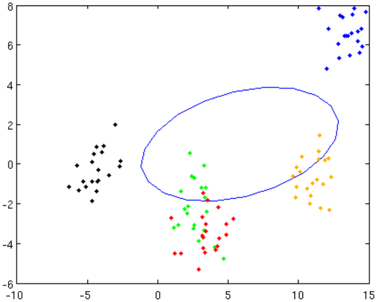


Model Selection with Dirichlet Processes

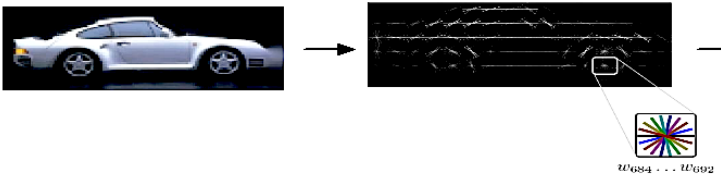
Goal: Algorithm defines by itself the optimal number of clusters without any prior knowledge

$\Downarrow$

Dirichlet Process Mixture Models



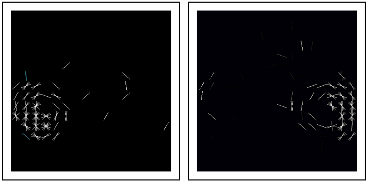
Application: Topic models



- **H**istogram of **O**riented **G**radient descriptors: Bag of Words Representation
- Topic: typical distribution of gradients
- Document: document specific distribution of topics
- Dirichlet Process finds the number of topics

topic model

$\Downarrow$

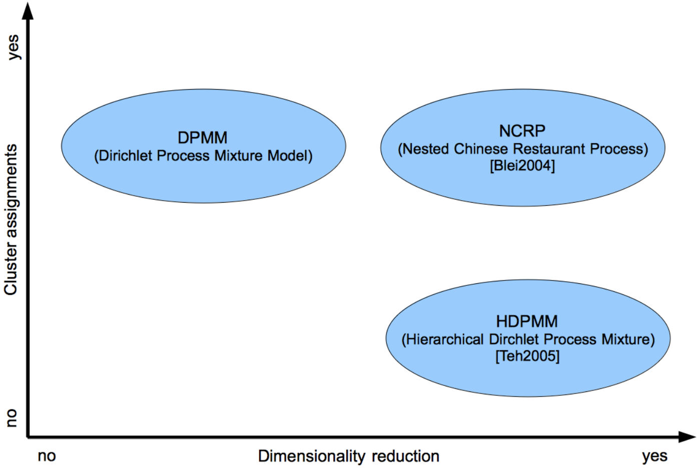


Open questions

- What are the properties of the Dirichlet Process?
- When does the process ’end’?
- Can one simply add new and unseen data?

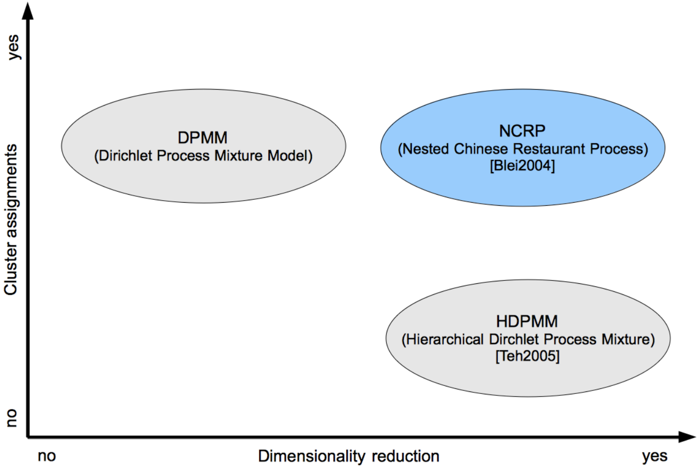
Dirichlet Processes in different models

Explore the Dirichlet Process in different models



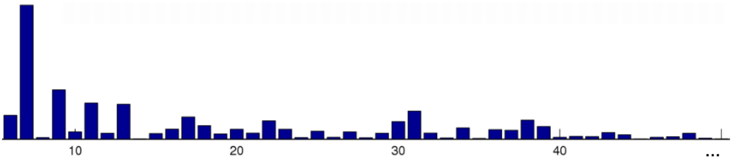
Dirichlet Processes in different models

Explore the Dirichlet Process in different models



Dirichlet Process

- Distribution of discrete distributions
- Possiblity of infinitely many bins



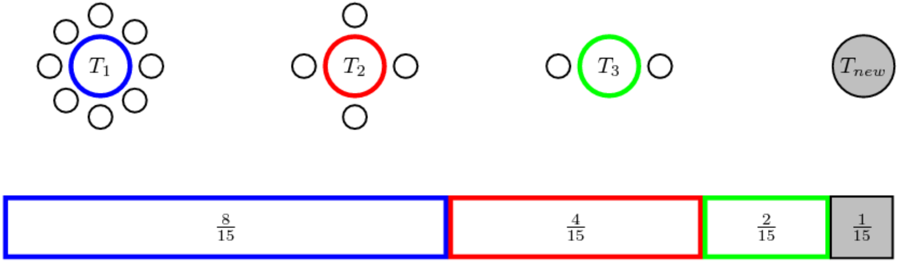
- Methods of representation:
  - Chinese Restaurant Process
  - Stick Breaking
  - Urne Scheme

Chinese Restaurant Process

Chinese Restaurant Process

Guest enters restaurant and decides on the basis of the current situation either for:

1. occupied table i: $\frac{m_i}{\gamma+m-1}$ (m_i guests on table i, m number of guests, γ parameter) or for
2. empty table: $\frac{\gamma}{\gamma+m-1}$

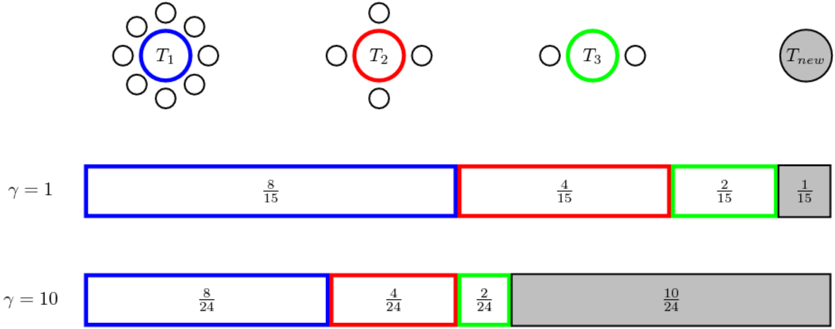


$\Rightarrow$ Clustering property

Chinese Restaurant Process

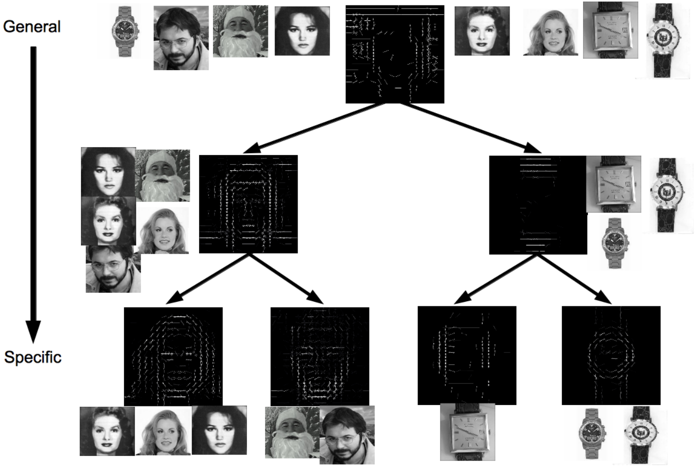
Guest enters restaurant and decides on the basis of the current situation either for:

1. occupied table i: $\frac{m_i}{\gamma+m-1}$ (m_i guests on table i, m number of guests, γ parameter) or for
2. empty table: $\frac{\gamma}{\gamma+m-1}$



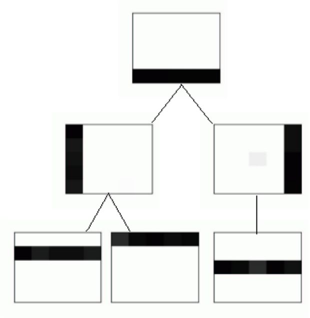
Nested Chinese Restaurant Processes [Blei2004]

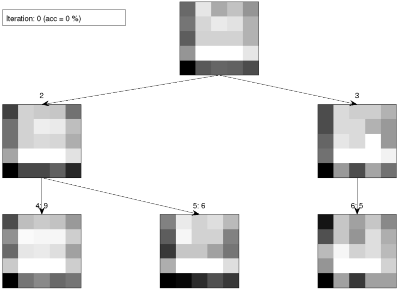
Each document is assigned to one path and is combined from the topics of the path.



Functioning of the NCRP on a toy example

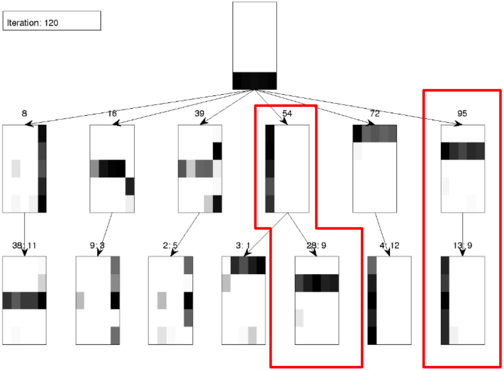
Recover the left tree with Nested Chinese Restaurant Process (NCRP)





Observation

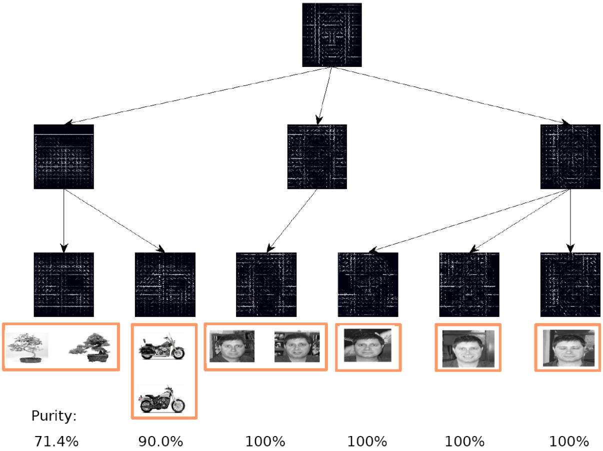
Paths with different arrangement of topics cannot be merged



Heuristic: the tokens are assigned randomly to the levels of the tree all n iterations $\Rightarrow$ decomposition of the fixed topic structure

Results: Hierarchy for Caltech 101 [Cao2007]

Partial hierarchy of NCRP with heuristic



Conclusion

- **What are the properties of the Dirichlet Process?** - clustering property means few clusters with many images and many clusters with only one or two images
- **When does the process 'end'?** - never, but you can observe the log-likelihood of the samples
- **Can you simply add new and unseen data?** - Yes, mostly after one ore two iterations

Thank you for your attention!