Non-parametric Bayesian Methods in Machine Learning

Dr. Simon Rogers
School of Computing Science
University of Glasgow
simon.rogers@glasgow.ac.uk
@sdrogers

May 9, 2014

Outline

- **FIX ME AT THE END**
- (My) Bayesian philosophy
- Gaussian Processes for Regression and Classification
  - GP preliminaries
  - Classification (including semi-supervised)
  - Regression application 1: clinical (dis)-agreement
  - Regressopn application 2: typing on touch-screens
- Dirichlet Process flavoured Cluster Models
  - DP preliminaries
  - Idenfitying metabolites
  - (if time) Cluster models for multiple data views

About me

- I'm not a statistican by training (don't ask me to prove anything!).
- Education:
  - Undergraduate Degree: Electrical and Electronic Engineering (Bristol)
  - PhD: Machine Learning Techniques for Microarray Analysis (Bristol)
- Currently:
  - Lecturer: Computing Science
  - Research Interests: Machine Learning and Applied Statistics in Computational Biology and Human-Computer Interaction (HCI)

Lecture 7: A mixture model for metabolite peak identification

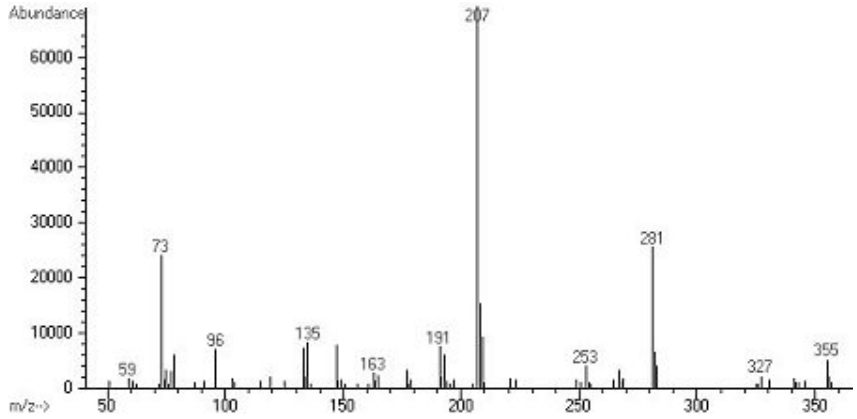
Dr. Simon Rogers
School of Computing Science
University of Glasgow
simon.rogers@glasgow.ac.uk
@sdrogers

May 9, 2014

Metabolomics

- Metabolome: the set of small molecule metabolites found within an organism.
  - Hormones, sugars, etc
- Gives a reliable picture of the phenotype (Fu et al 2009)
- But metabolites are hard to measure.
- Dominant paradigm is Liquid Chromatography (LC) Mass Spectrometry (MS)

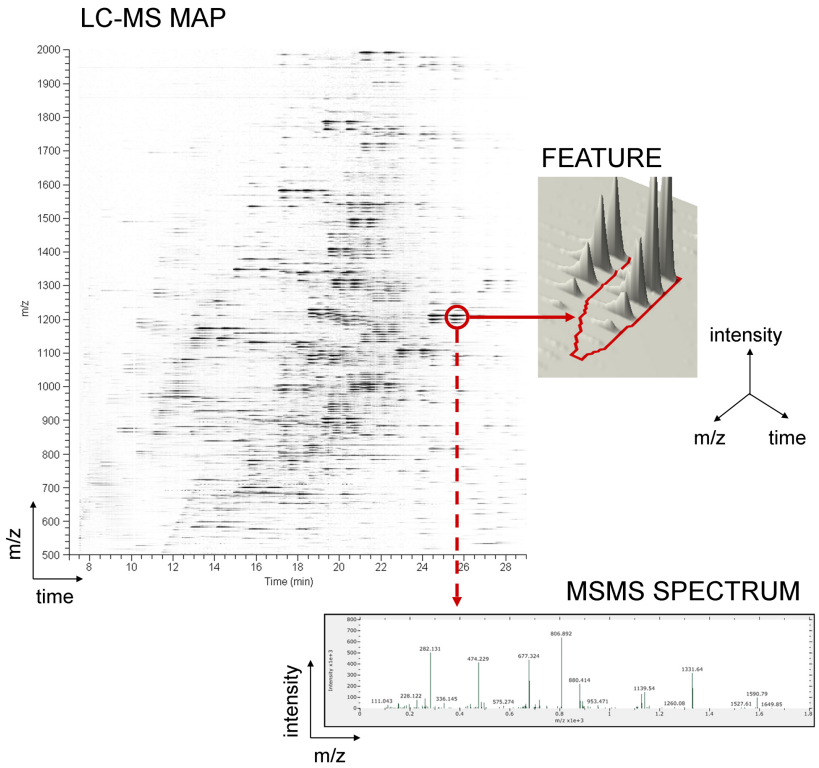
MS



- Output of MS is a set of mass-intensity pairs (peaks).
- Each peak corresponds to one ion.
- Each metabolite can result in many different ions:
  - Different ions (i.e. H^+, K^+)
  - Isotopes
- All have predictable theoretical mass (for particular metabolite)

LC/MS

- Most samples are too complex for a single MS analysis
- First *separate* the sample via LC
- Perform many MS analysis at different Retention Time (RT)

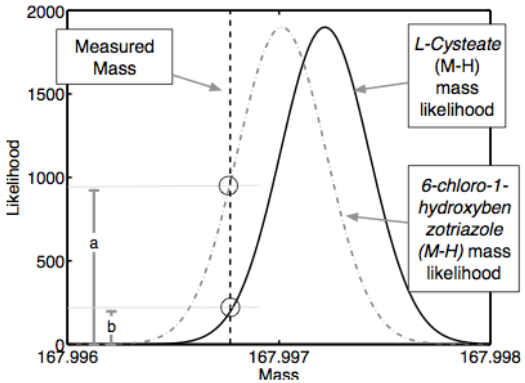


(image from Peltoniemi et. al)

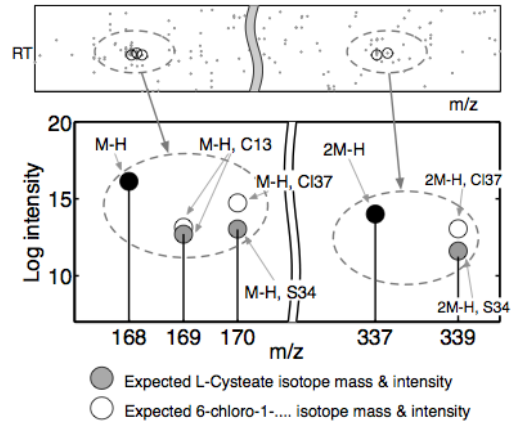
What are the peaks?

- How do we identify things in this 2D image?
- Doing each peak separately (traditional approach; searching mass against a database) leads to many false positives
- Can we use the fact that all peaks for a single metabolite will have very similar RT?
  - If we look for a particular metabolite, we should see a predictable set of mass peaks at the same RT
  - Use a Dirichlet Process (DP) mixture model...
  - Using the dependency information between peaks to improve identification
  - Get probabilities of identification rather than hard decisions

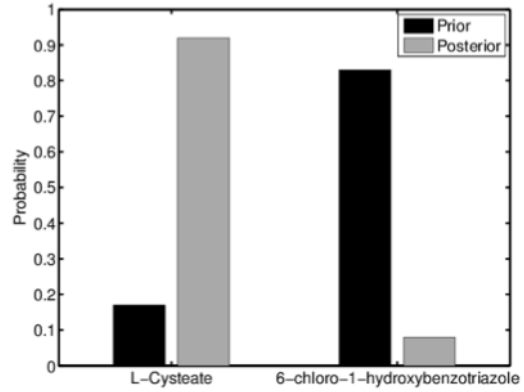
MetAssign



(a) Relative mass likelihoods for two different formulas



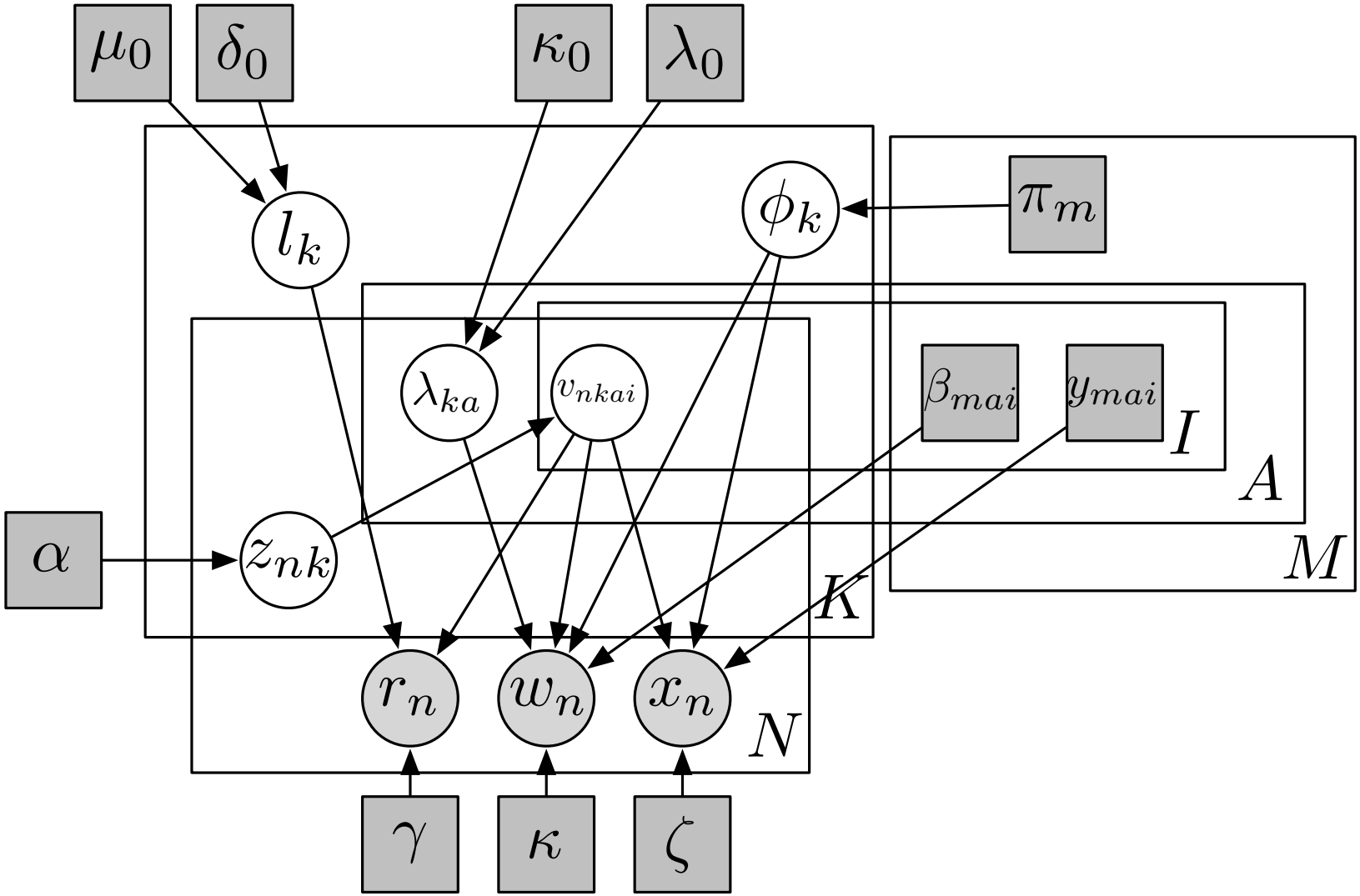
(b) Peaks with similar rentention time included in the cluster with the peak at m/z=168. The circles show the expected intensity values of isotope peaks



(c) The change in probability from the prior to the posterior

- Database matches for each peak can be thought of as prior annotations
- After clustering we have posterior matches
  - Note that by averaging over all clusterings we get posterior assignments for each peak.
  - i.e. we are not interested in one clustering.

MetAssign


μ_0
δ_0
κ_0
λ_0
l_k
ϕ_k
π_m
λ_{ka}
v_{nkai}
β_{mai}
y_{mai}
I
A
M
α
z_{nk}
K
r_n
w_n
x_n
N
γ
κ
ζ

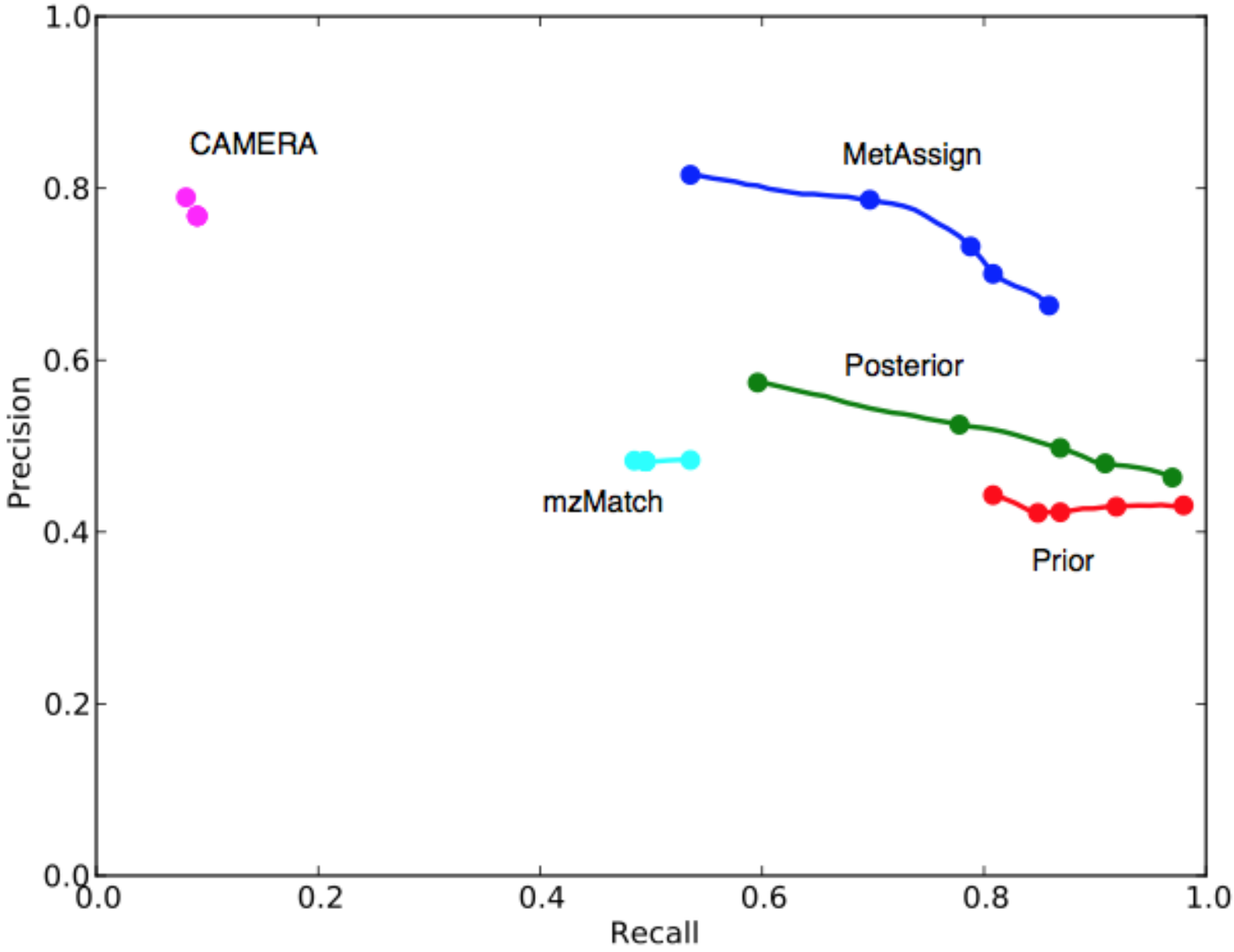
MetAssign

- Model consists of K clusters
- Each cluster is linked to a metabolite from the database (ϕ_k)
- The metabolite links lets us work out what masses (y_{mai}) and intensity relationships (β_{mai}) (for isotopes) we ought to see
- Each cluster has a retention time l_k
- Each adduct (ionisation type) has an intensity (λ_{ka})
- z_{nk} defines cluster membership for peak n
- v_{nkai} defines membership within the cluster

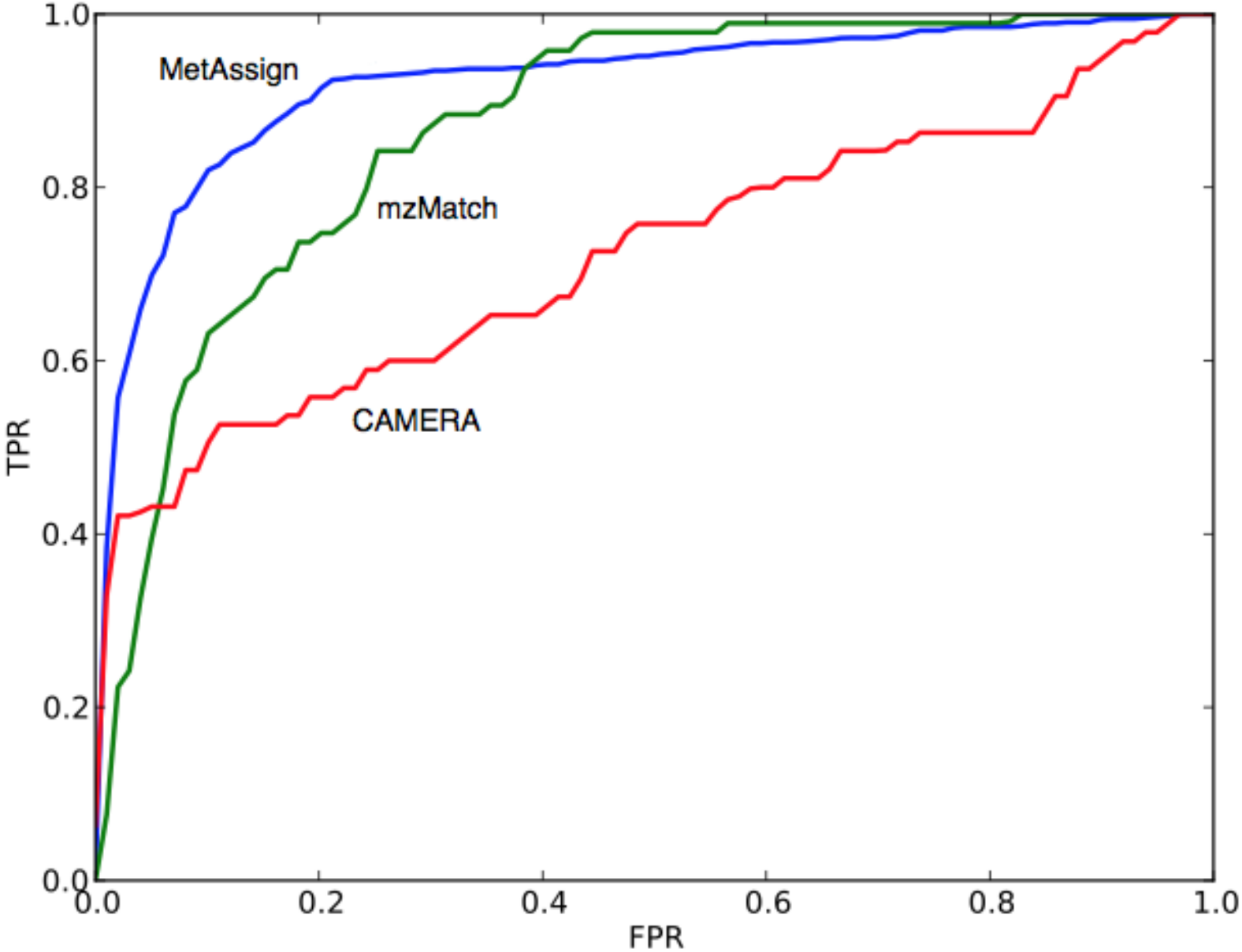
MetAssign: inference

- Gibbs sampling updates are all fairly straightforward (assuming Gaussian noise everywhere)
- Can also include Metropolis-Hastings steps for changing which metabolite a cluster is assigned to
- Quite slow
- Proportion of times a peak is assigned to a metabolite (via a cluster) gives posterior probability
- Access to posterior samples lets us do useful things
  - e.g. only consider assignments of peaks if all bigger isotope peaks are present

MetAssign: results



MetAssign: results



Conclusions

- Excellent performance (better than state of the art)
- DP prior allows us to not fix number of metabolites *a-priori*
- Probabilities are obtained by averaging over the clusterings
- Probabilistic assignments are useful for the experimenters
- Gibbs sampling is easy to implement (although a pain to make efficient)

Future work

- Incorporate predicted RT
  - It is possible to predict (badly) the RT of a particular metabolite
  - Can easily be incorporated into the model through a metabolite-specific RT prior

Future work

- Incorporate predicted RT
  - It is possible to predict (badly) the RT of a particular metabolite
  - Can easily be incorporated into the model through a metabolite-specific RT prior
- Incorporate connectivity
  - Metabolites live in networks (e.g. reactions)
  - We've previously used this to aid identification (Rogers et. al 2009)
  - Can be incorporated into the Metropolis-Hastings step to *a-priori* make highly connected metabolites more likely

Future work

- Incorporate predicted RT
  - It is possible to predict (badly) the RT of a particular metabolite
  - Can easily be incorporated into the model through a metabolite-specific RT prior
- Incorporate connectivity
  - Metabolites live in networks (e.g. reactions)
  - We've previously used this to aid identification (Rogers et. al 2009)
  - Can be incorporated into the Metropolis-Hastings step to *a-priori* make highly connected metabolites more likely
- Note: MetAssign currently under revision. When (if) published, search for 'Rogers Daly Breitling metassign bioinformatics'

Lecture 8: The Hierarchical Dirichlet Process

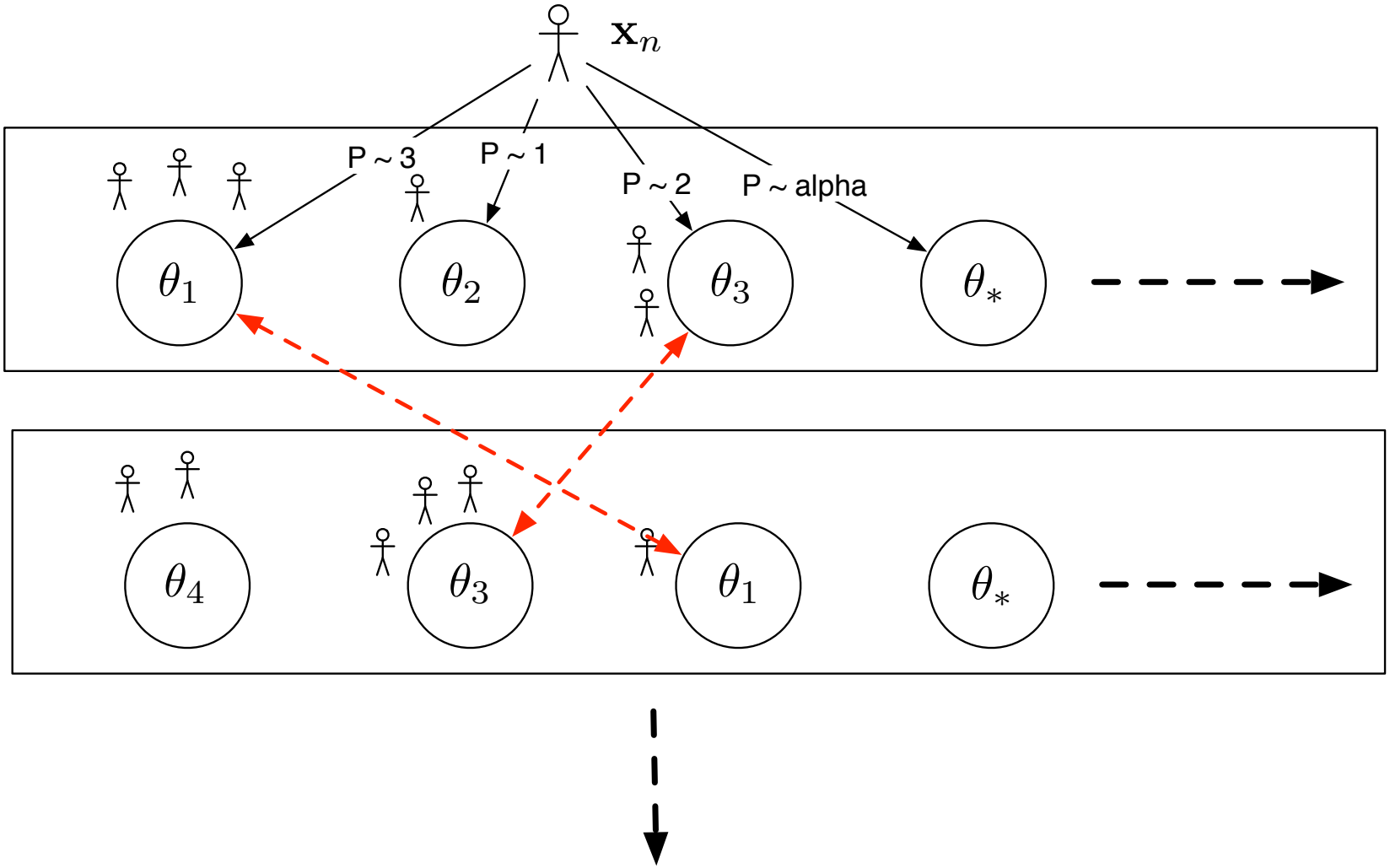
Dr. Simon Rogers
School of Computing Science
University of Glasgow
simon.rogers@glasgow.ac.uk
@sdrogers

May 9, 2014

The Hierarchical DP

- Imagine we have > 1 related dataset to cluster, generated by the same process
- Fitting separate mixtures to each results in a loss of information
- The Hierarchical Dirichlet Process (HDP) allows them to be analysed together with shared parameters
  - e.g. datasets clustered individually but cluster parameters (e.g. means) can be *shared*
  - Analogy: The Chinese Restaurant Franchise
  - Described in Teh et. al

The Chinese Restaurant Franchise



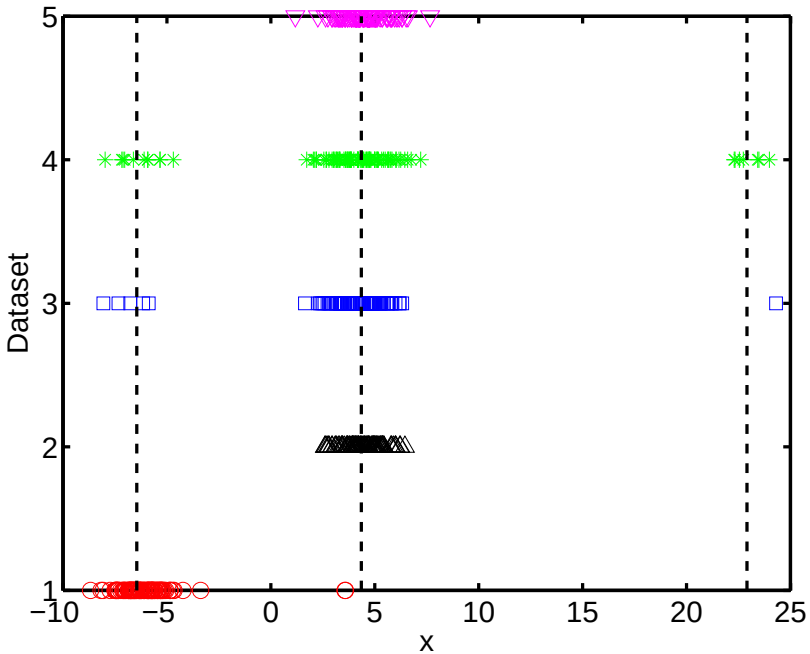
The Chinese Restaurant Franchise

- Gibbs sampling is very similar to the single restaurant case
- Each object is re-assigned to a table with probabilities (i indexes datasets):

$$P(z_{ink} = 1 | \ldots) \propto \begin{cases} n_{ik} p(\mathbf{x}_{in} | \theta_k) & \text{for current table} \\ \alpha p(\mathbf{x}_{in}) & \text{for new } k \end{cases}$$

- $p(\mathbf{x}_{in})$ is computed by marginalising all possible values for the parameters at the new tables.
  - These include values used for current tables (with prior proportional to the number of tables they're used at) and a completely new value (with prior proportional to, say, γ).
  - i.e. There are DPs for assignment of objects to tables and assignment tables to parameters.

Data sampled from a HDP



- 5 datasets
- 4 top level components (dashed line)
- Each data has data from a subset