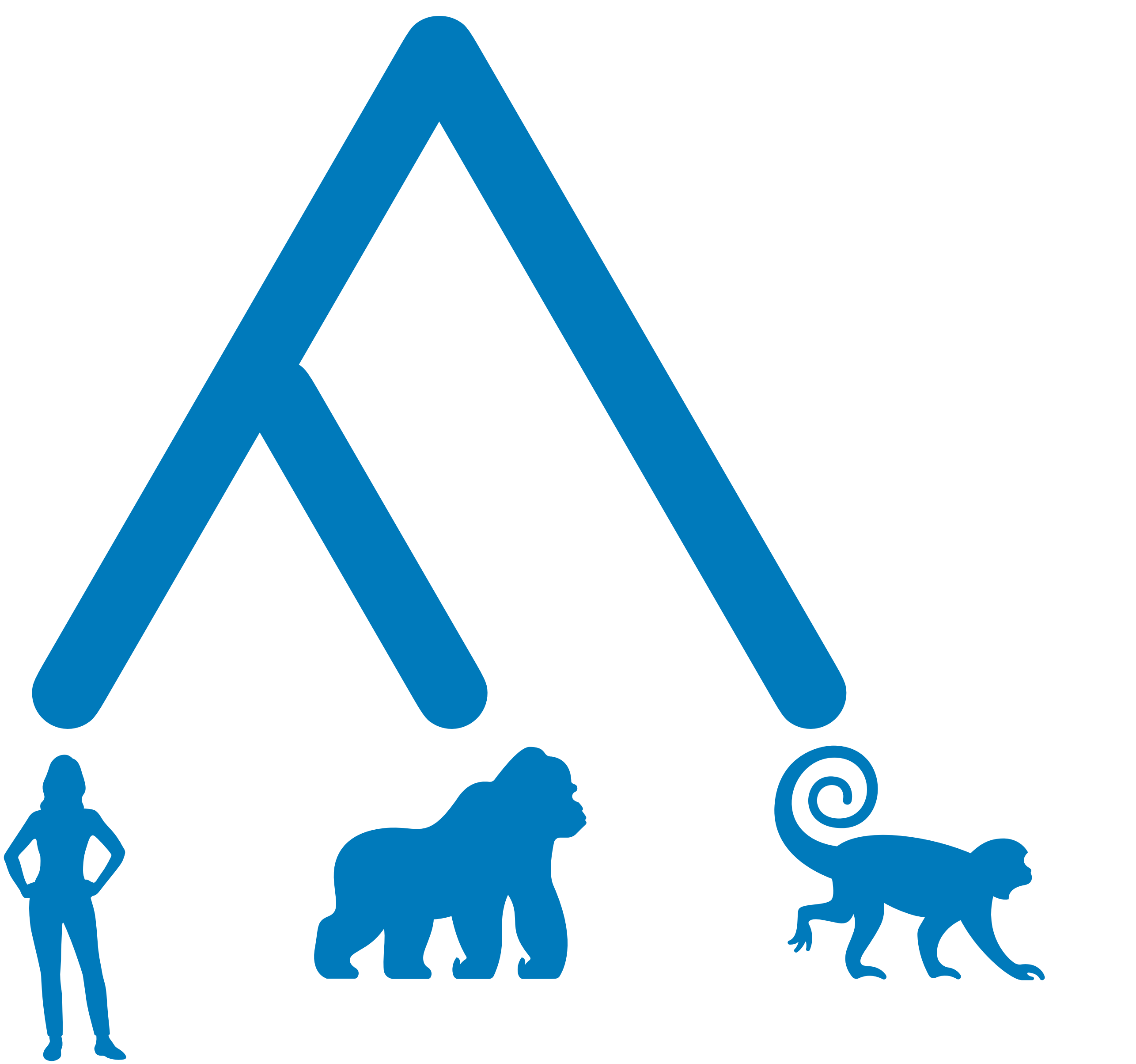
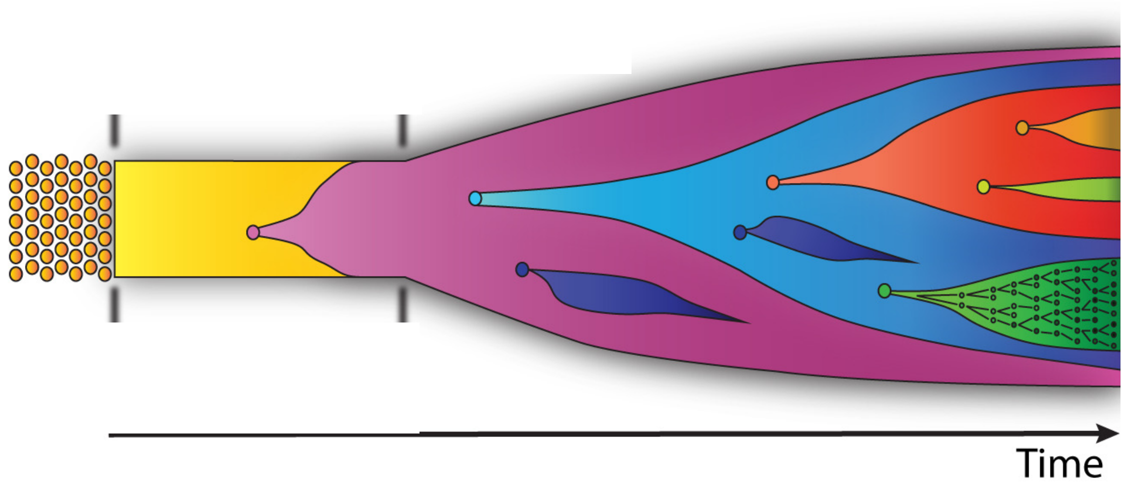
Royal Institute of Technology

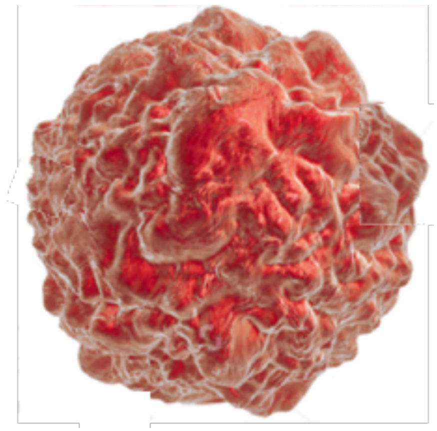
DD2434 STAT. ADVANCE MACHINE LEARNING HT 2017

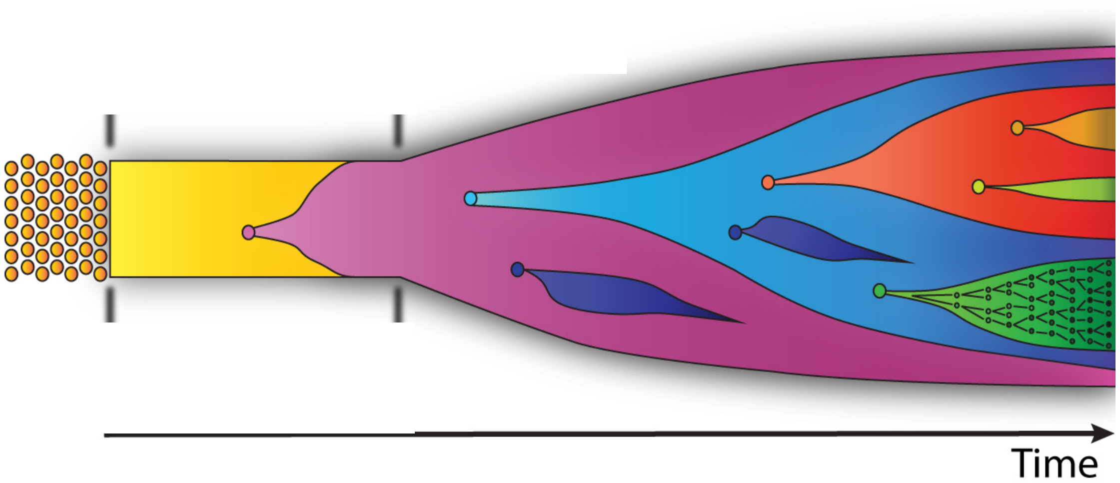
★ Lecture 1-Intro

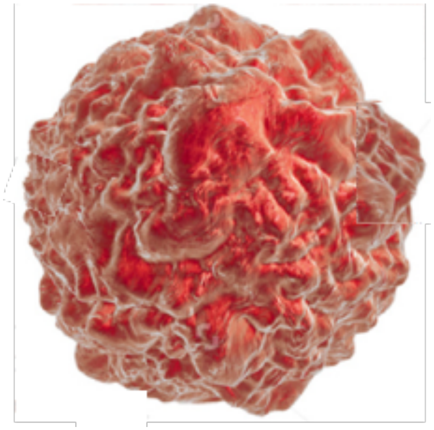
"*Chapter 1*"

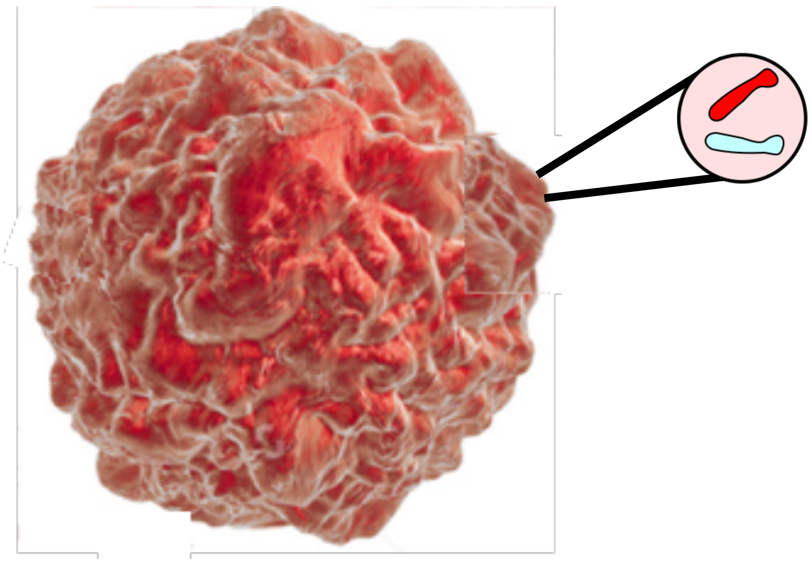


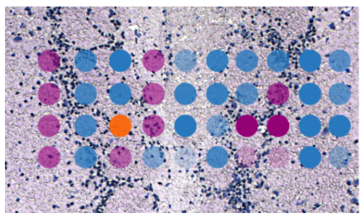

Time

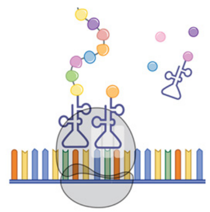


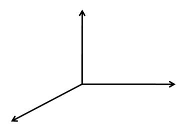

Time

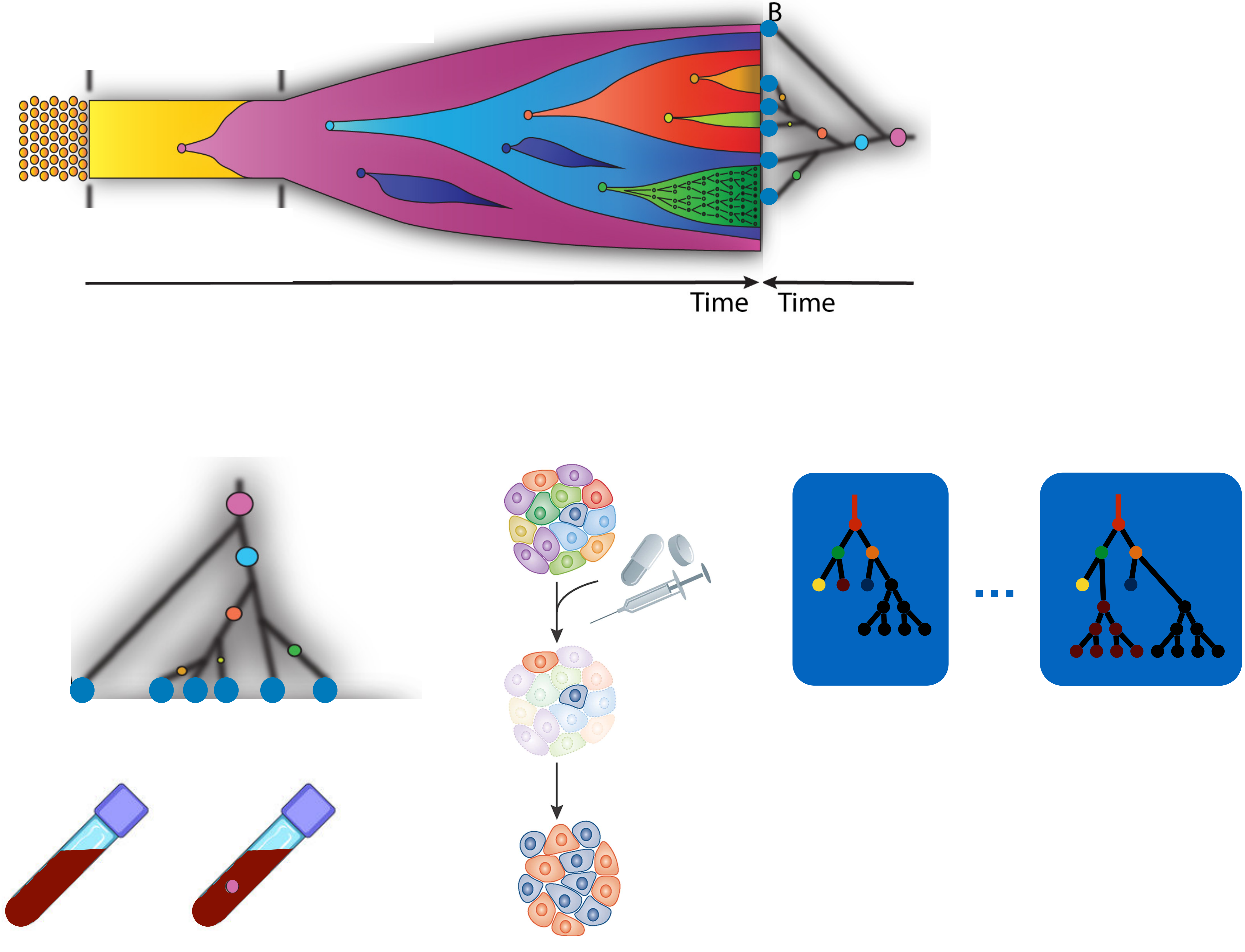


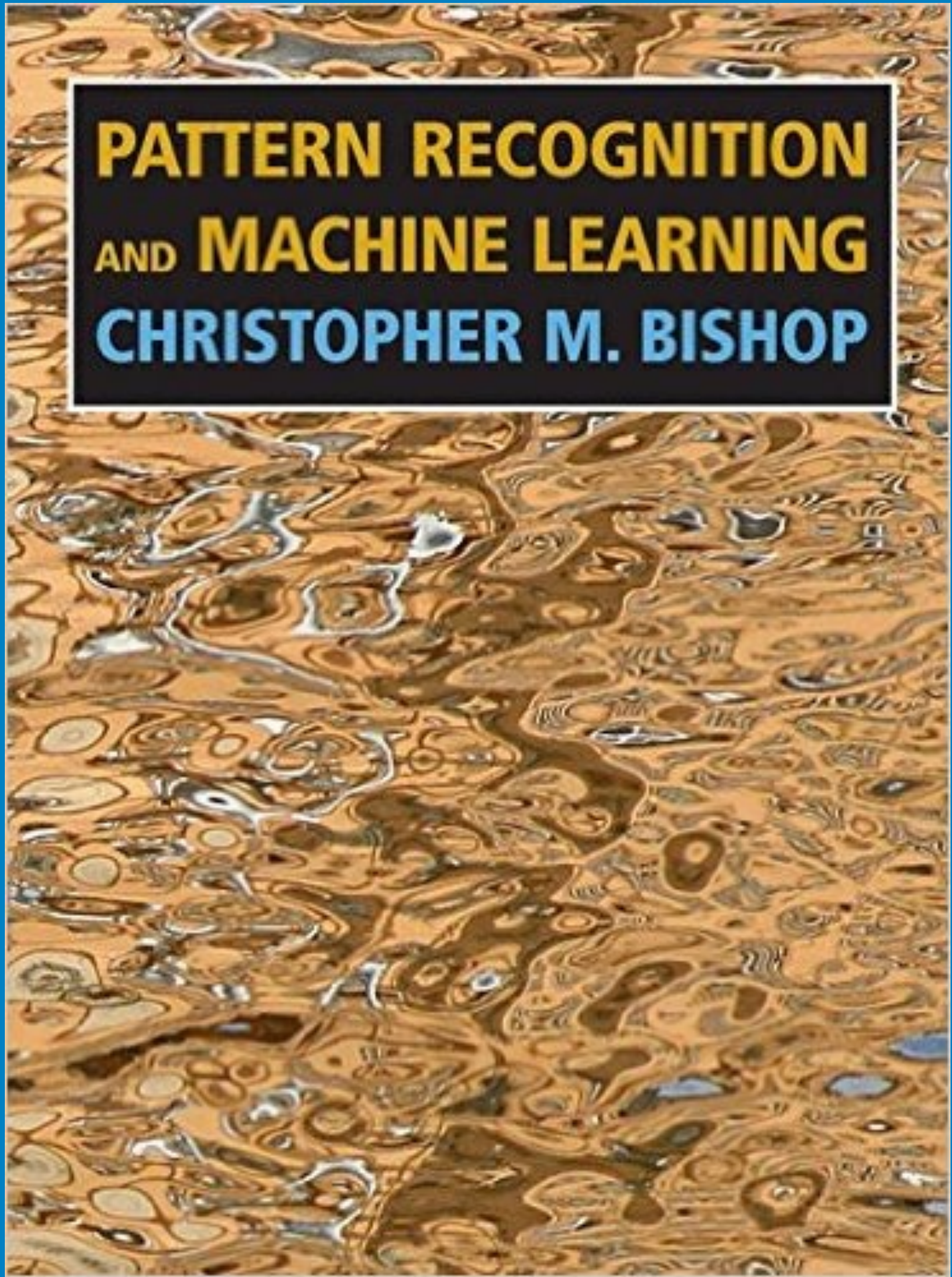









B
Time
Time



THE TEXT BOOK

- Key machine learning researcher
- Reasonably mathematical
- Well-written

CHANGES

- Videos and Q & A
- Assignments and project
- Not Gaussian Processes
- More on prior-posterior
- New coverage of dimensionality reduction
- New teacher Aristides Gionis
  - Data science
  - Athens, Stanford University,Yahoo! Research, Aalto University

ADVICE FROM EARLIER YEARS

- Read Bishop's book
- Go to all lectures. Read up. Start with the project on time.
- Avoid the lectures, study Bishop from the beginning because in the other hand you will find yourself in a position that you will be running to catch up with the deadlines.

ADVICE & COMMENTS

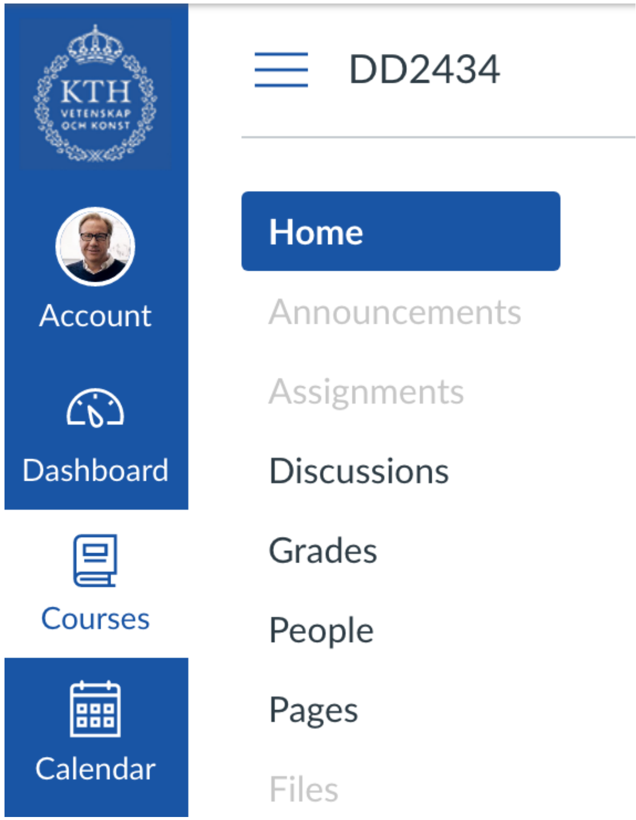
- The assignments were thorough, so you actually had to understand the material to pass. The course contents were very interesting.
- The assignments takes at least three times longer than you think, start on time with everything.
- Don't underestimate the time it takes to finish the assignments Do the project and assignment as early as possible
- The problems are intentionally made so you cannot Google them…
- In addition, the lectures did not reflect at all on what we should go through in the assignments, so it was basically a matter of own work…
- Solving the assignments made me understand and made everything clear…

or

- ★ Assignments (3) – Individual, handed in and graded
  - ★ First out now, or next week, in December 1(D,E)
  - ★ Second out November 15, in Dec 15 (D,E)
  - ★ Third out November 25, in Jan 15 (A,B,C)

- ★ Project (1) – Collaborative, handed in and graded
  - ★ Out December 1, in Jan 15 (A-C, D-E)
  - ★ You define the groups of four (based on merit)

APPROXIMATE DATES — EXAMINATION



INTERACTION

- Lectures, exercises Zoom
- Solutions: CANVAS.
- Slides, notes, videos: Canvas
- Assignments, projects: Slack, Zoom sessions

1. Use latex, you will have to use it later anyway (this is a recommendation)

2. Always include your names in the file with the solutions

3. Make each step in a derivation explicit

CONCERNING SOLUTIONS

MAKING SENSE OF DATA

ARTICLE

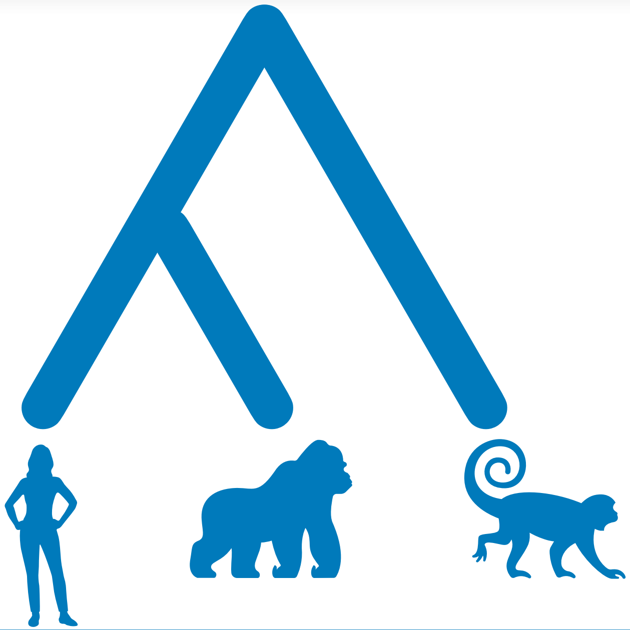
https://doi.org/10.1038/s42003-020-01247-y OPEN

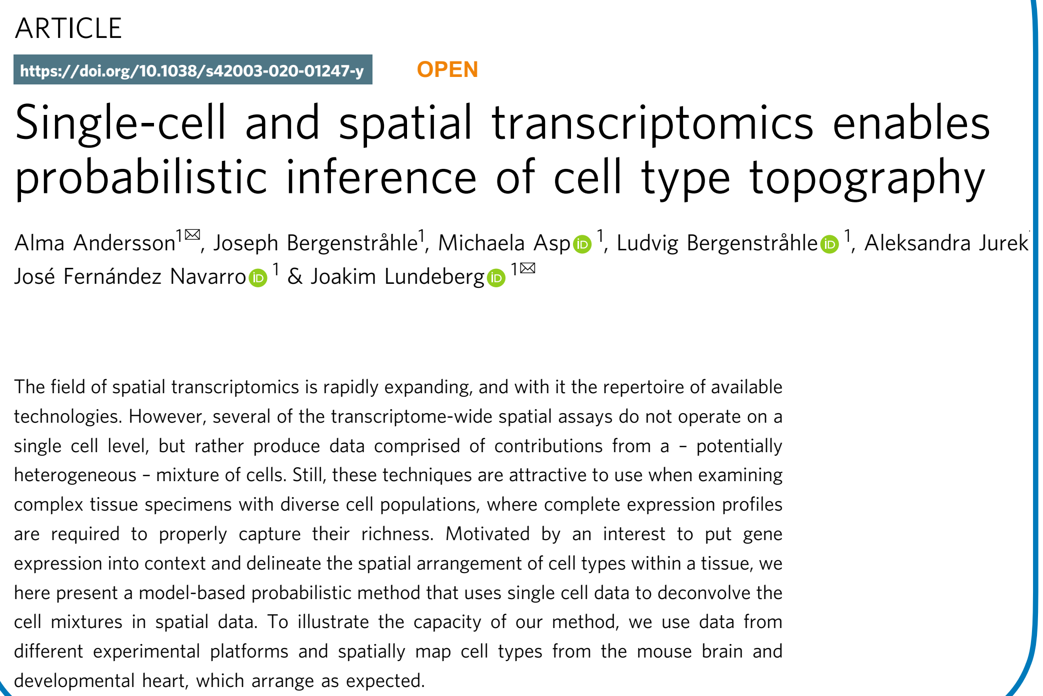
Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography

Alma Andersson[1✉], Joseph Bergenstråhle[1], Michaela Asp[1], Ludvig Bergenstråhle[1], Aleksandra Jurek
José Fernández Navarro[1] & Joakim Lundeberg[1✉]

The field of spatial transcriptomics is rapidly expanding, and with it the repertoire of available technologies. However, several of the transcriptome-wide spatial assays do not operate on a single cell level, but rather produce data comprised of contributions from a – potentially heterogeneous – mixture of cells. Still, these techniques are attractive to use when examining complex tissue specimens with diverse cell populations, where complete expression profiles are required to properly capture their richness. Motivated by an interest to put gene expression into context and delineate the spatial arrangement of cell types within a tissue, we here present a model-based probabilistic method that uses single cell data to deconvolve the cell mixtures in spatial data. To illustrate the capacity of our method, we use data from different experimental platforms and spatially map cell types from the mouse brain and developmental heart, which arrange as expected.

PROBABILISTIC APPROACH: EXPLANATION - LATENT VARIABLE





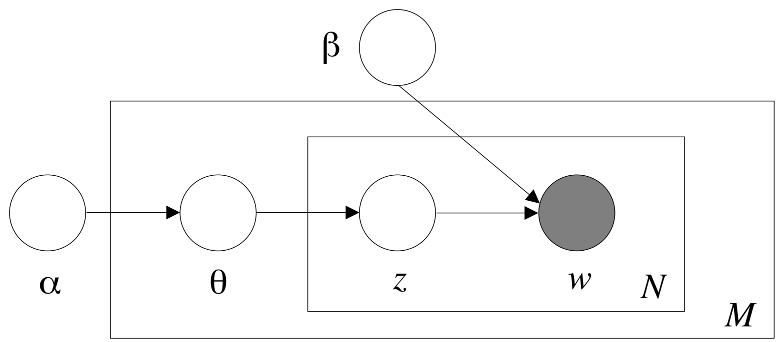
ARTICLE

https://doi.org/10.1038/s42003-020-01247-y OPEN

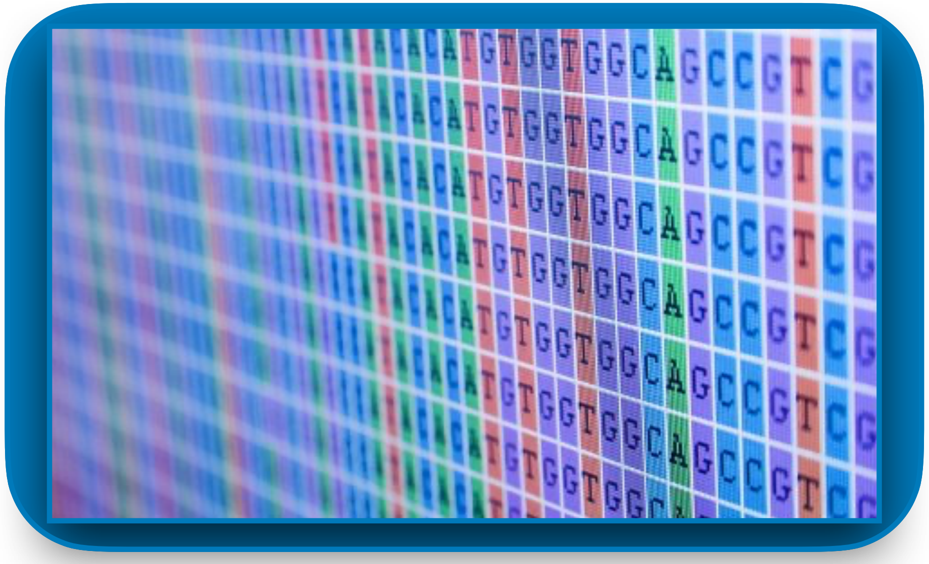
Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography

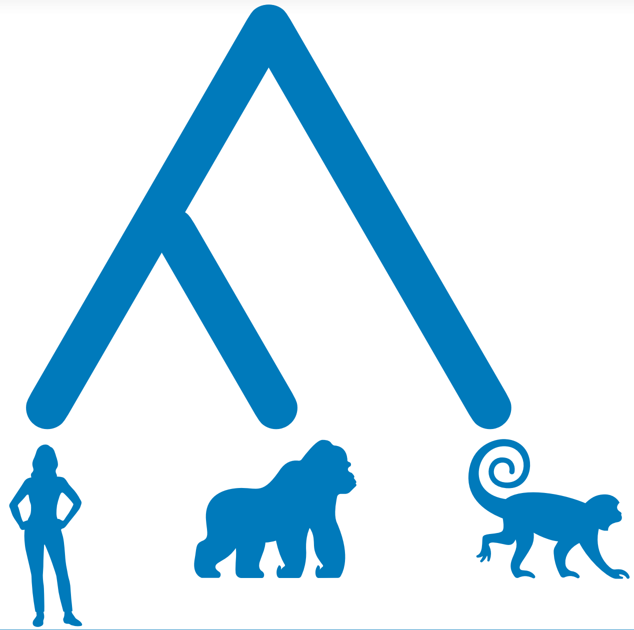
Alma Andersson[1]✉, Joseph Bergenstråhle[1], Michaela Asp[1], Ludvig Bergenstråhle[1], Aleksandra Jurek
José Fernández Navarro[1] & Joakim Lundeberg[1]✉

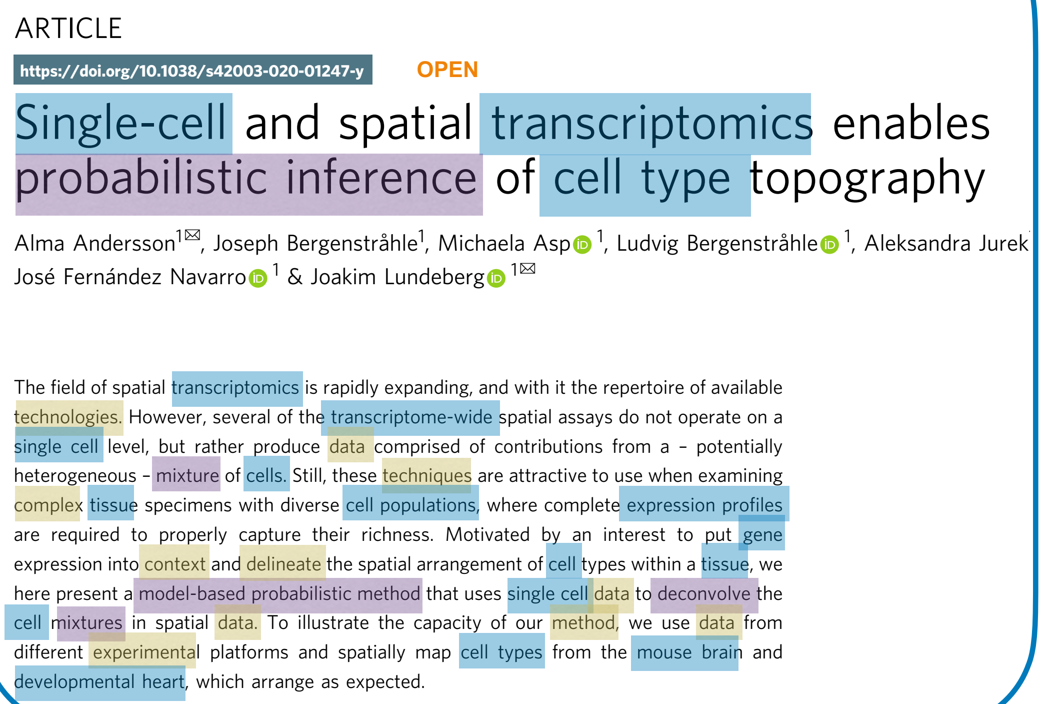
The field of spatial transcriptomics is rapidly expanding, and with it the repertoire of available technologies. However, several of the transcriptome-wide spatial assays do not operate on a single cell level, but rather produce data comprised of contributions from a – potentially heterogeneous – mixture of cells. Still, these techniques are attractive to use when examining complex tissue specimens with diverse cell populations, where complete expression profiles are required to properly capture their richness. Motivated by an interest to put gene expression into context and delineate the spatial arrangement of cell types within a tissue, we here present a model-based probabilistic method that uses single cell data to deconvolve the cell mixtures in spatial data. To illustrate the capacity of our method, we use data from different experimental platforms and spatially map cell types from the mouse brain and developmental heart, which arrange as expected.



PROBABILISTIC APPROACH: EXPLANATION - LATENT VARIABLE







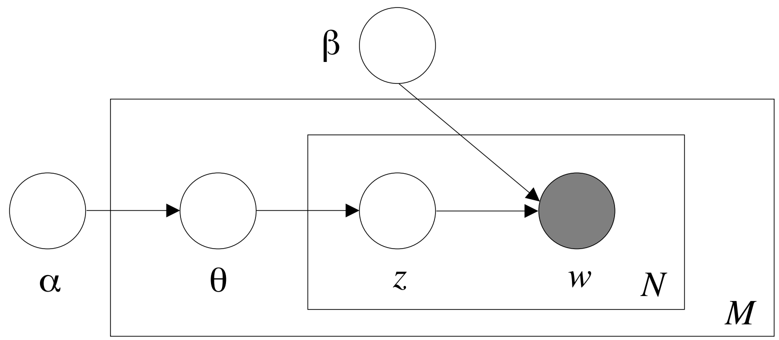
ARTICLE

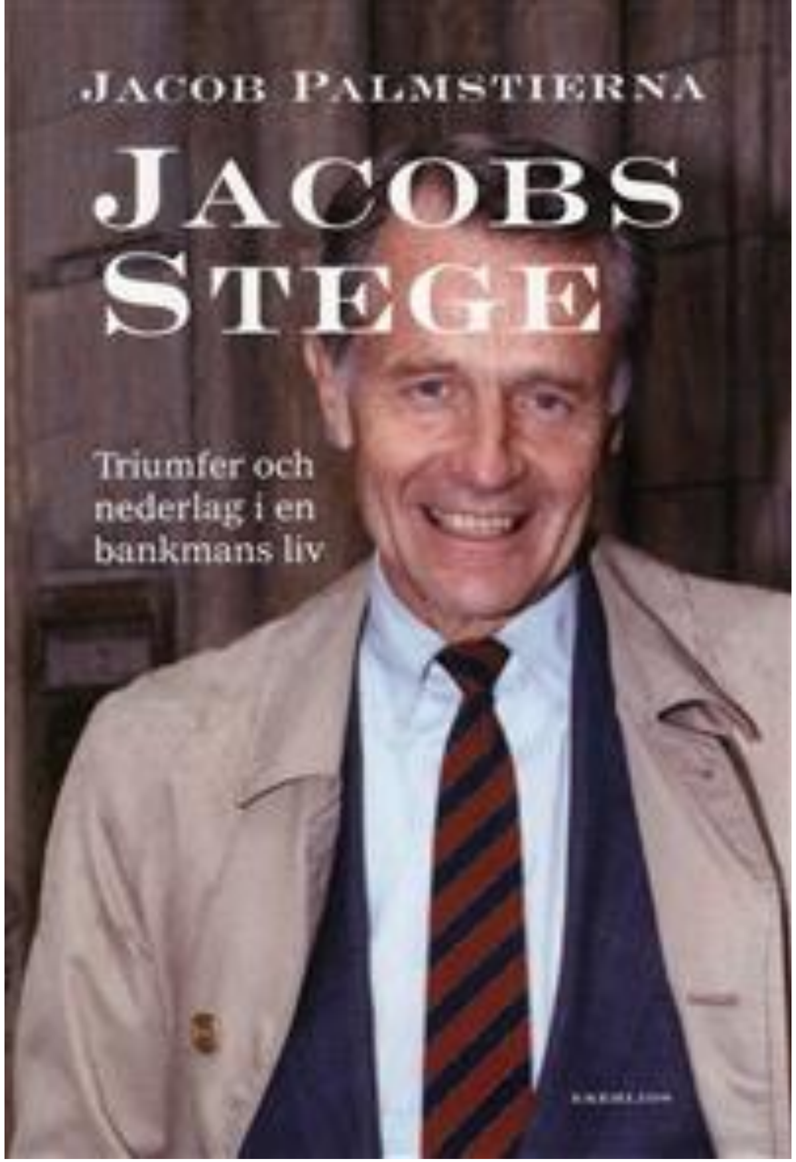
https://doi.org/10.1038/s42003-020-01247-y OPEN

Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography

Alma Andersson[1✉], Joseph Bergenstråhle[1], Michaela Asp[1], Ludvig Bergenstråhle[1], Aleksandra Jurek
José Fernández Navarro[1] & Joakim Lundeberg[1✉]

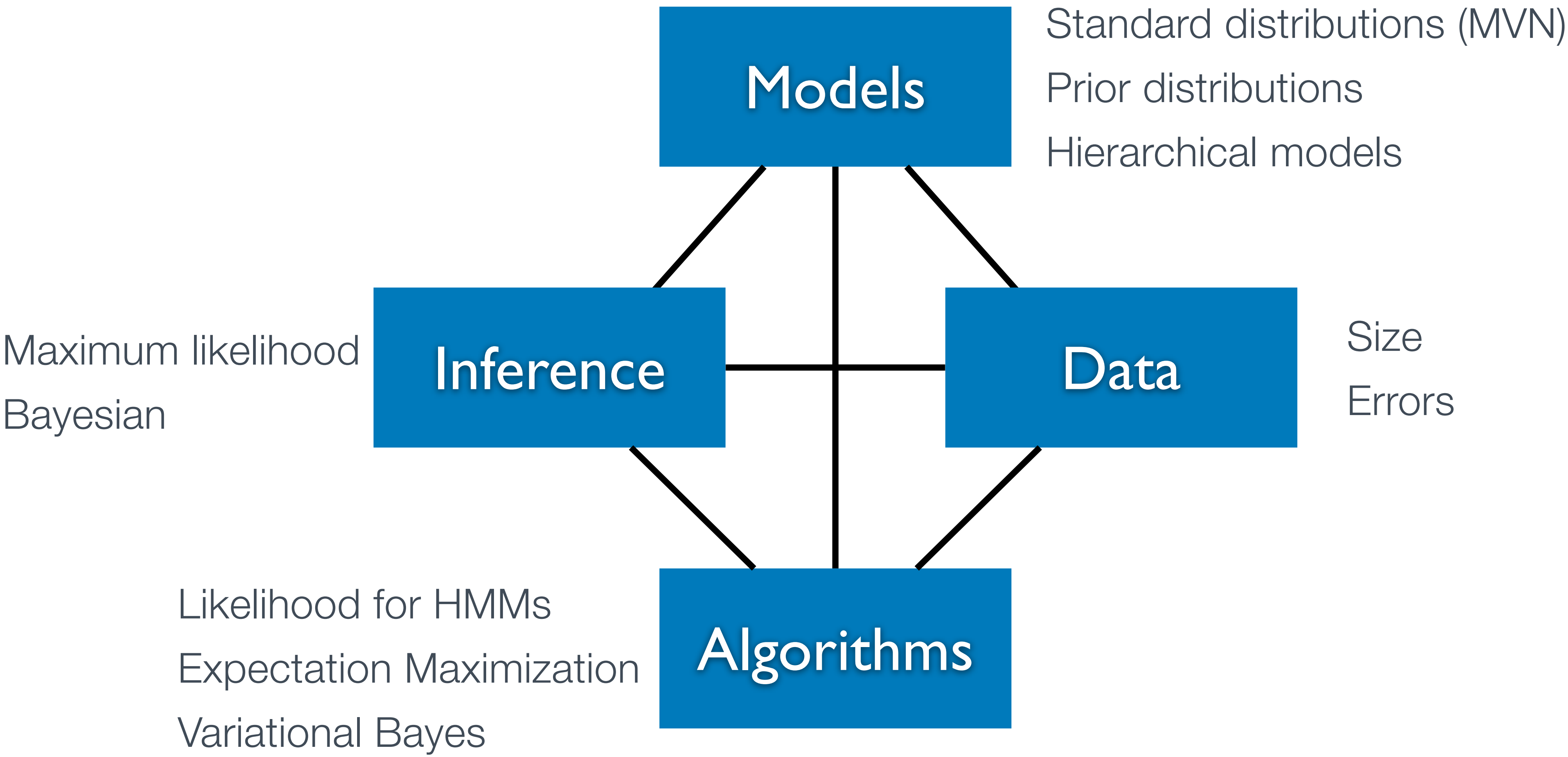
The field of spatial transcriptomics is rapidly expanding, and with it the repertoire of available technologies. However, several of the transcriptome-wide spatial assays do not operate on a single cell level, but rather produce data comprised of contributions from a – potentially heterogeneous – mixture of cells. Still, these techniques are attractive to use when examining complex tissue specimens with diverse cell populations, where complete expression profiles are required to properly capture their richness. Motivated by an interest to put gene expression into context and delineate the spatial arrangement of cell types within a tissue, we here present a model-based probabilistic method that uses single cell data to deconvolve the cell mixtures in spatial data. To illustrate the capacity of our method, we use data from different experimental platforms and spatially map cell types from the mouse brain and developmental heart, which arrange as expected.





TAKING DECISIONS IN THE PRESENCE OF UNCERTAINTY

PROBABILISTIC MACHINE LEARNING



A STOCHASTIC APPROXIMATION METHOD[1]

By Herbert Robbins and Sutton Monro

University of North Carolina

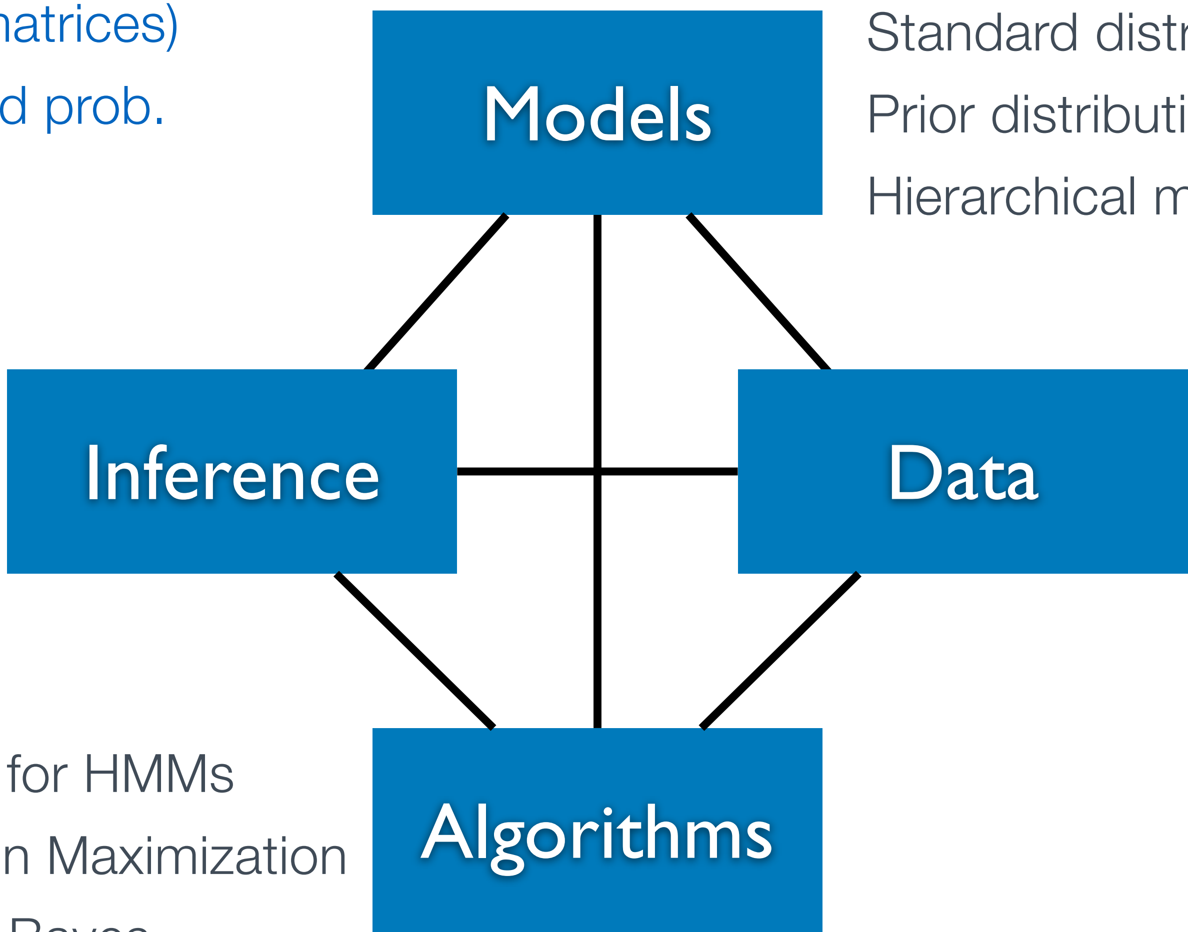
1. Summary. Let $M(x)$ denote the expected value at level x of the response to a certain experiment. $M(x)$ is assumed to be a monotone function of x but is unknown to the experimenter, and it is desired to find the solution $x = \theta$ of the equation $M(x) = \alpha$, where α is a given constant. We give a method for making successive experiments at levels $x_1, x_2, \cdots$ in such a way that x_n will tend to θ in probability.

The Annals of Mathematical Statistics
Vol. 22, No. 3 (Sep., 1951), pp. 400-407 (8 pages)
Published By: Institute of Mathematical Statistics

PROBABILISTIC MACHINE LEARNING

Have a look at

- Linear algebra (matrices)
- Previous stat. and prob. courses
- Chapter 2.



Standard distributions (MVN)
Prior distributions
Hierarchical models

Maximum likelihood
Bayesian

Size
Errors

Likelihood for HMMs
Expectation Maximization
Variational Bayes

MACHINE LEARNING
DD2447

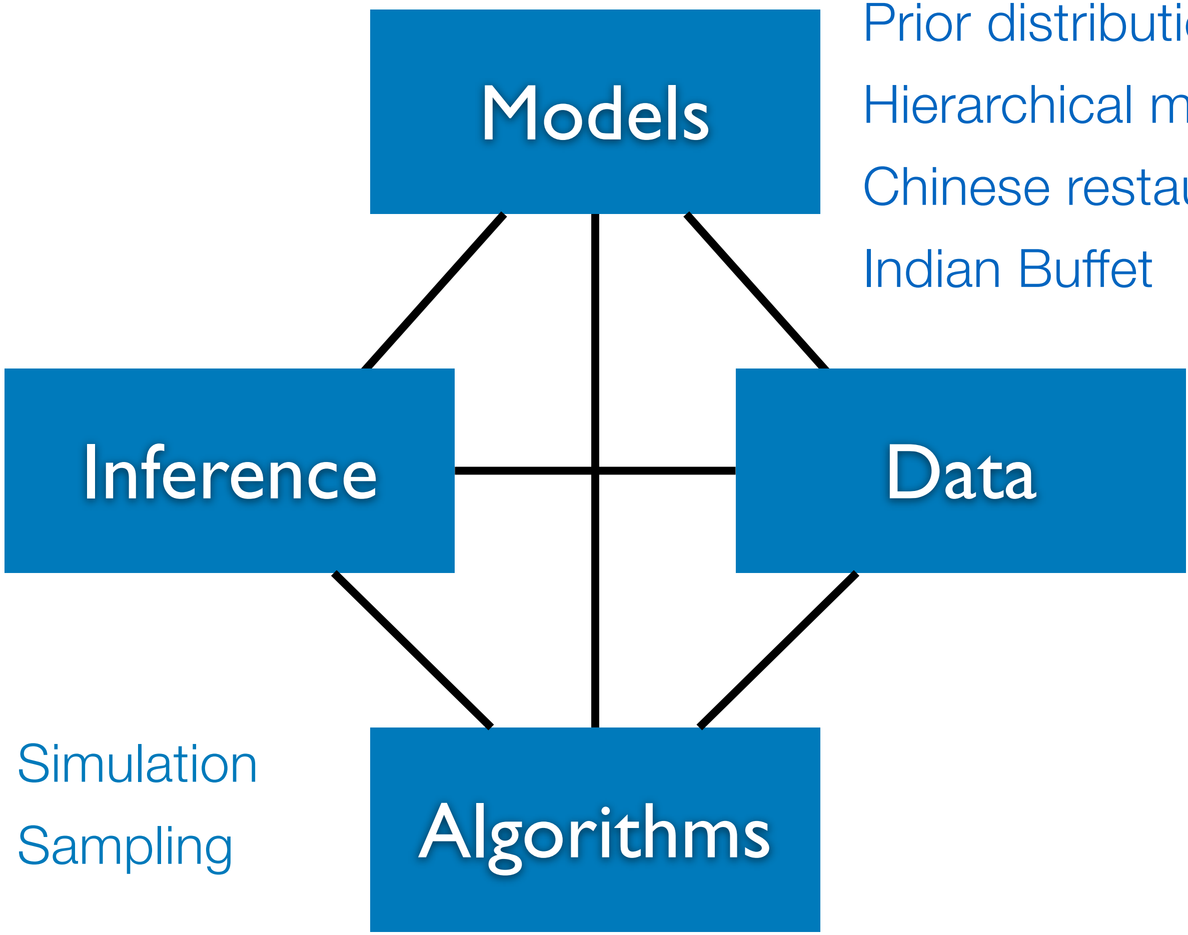
Standard distributions (MVN)

Prior distributions

Hierarchical models

Chinese restaurant

Indian Buffet



Maximum likelihood

Bayesian

Size

Errors

Simulation

Sampling

SOME THOUGHTS ON MODELING

* All models are wrong, but some are useful.
  * Models are what we call the lies we are used to
* There are no model free approaches!
  * use the term assumption instead
* Using models is a way to make assumptions explicit.
* Bayesian is a non-deterministic logic.

SOME STUFF I EXPECT YOU TO KNOW

- Supervised learning
- Unsupervised learning
- Training & testing
- Probability

MAXIMUM LIKELIHOOD (ML) AND POSTERIOR PREDICTIVE

- ML
  - Estimate θ_{ML} from training data D and then use

$$y(x, \theta_{ML})$$

- Bayesian
  - Estimate a posterior distribution over θ based on D and then use

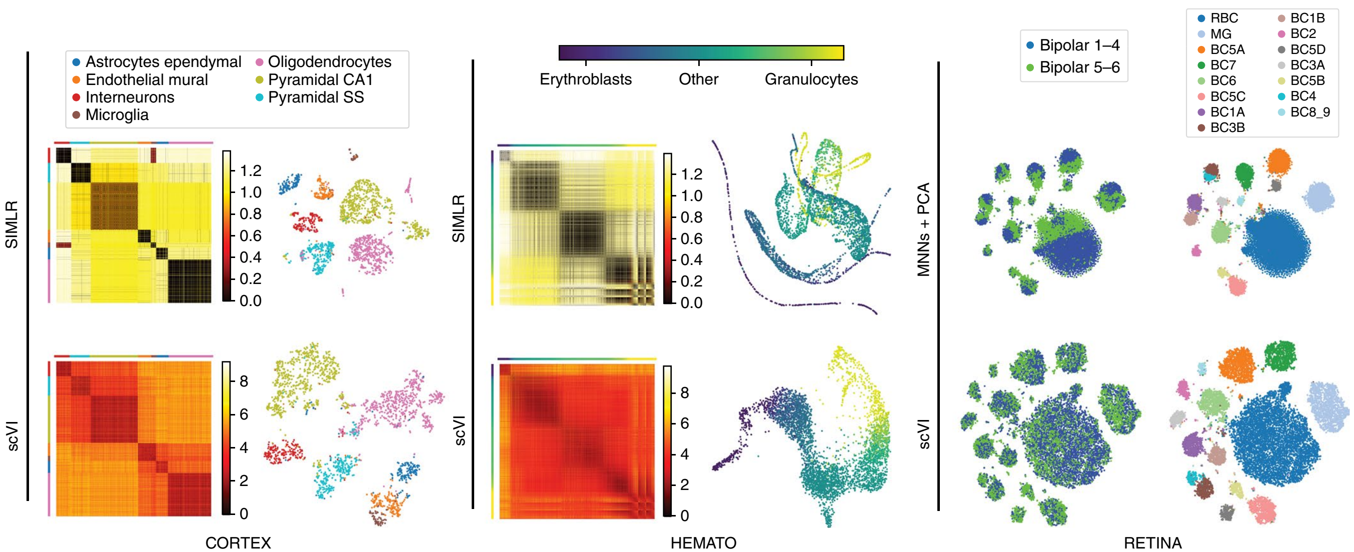
$$\int y(x, \theta) p(\theta \mid \mathscr{D}) d\theta$$

DIMENSIONALITY REDUCTION - VARIATIONAL AUTOENCODER

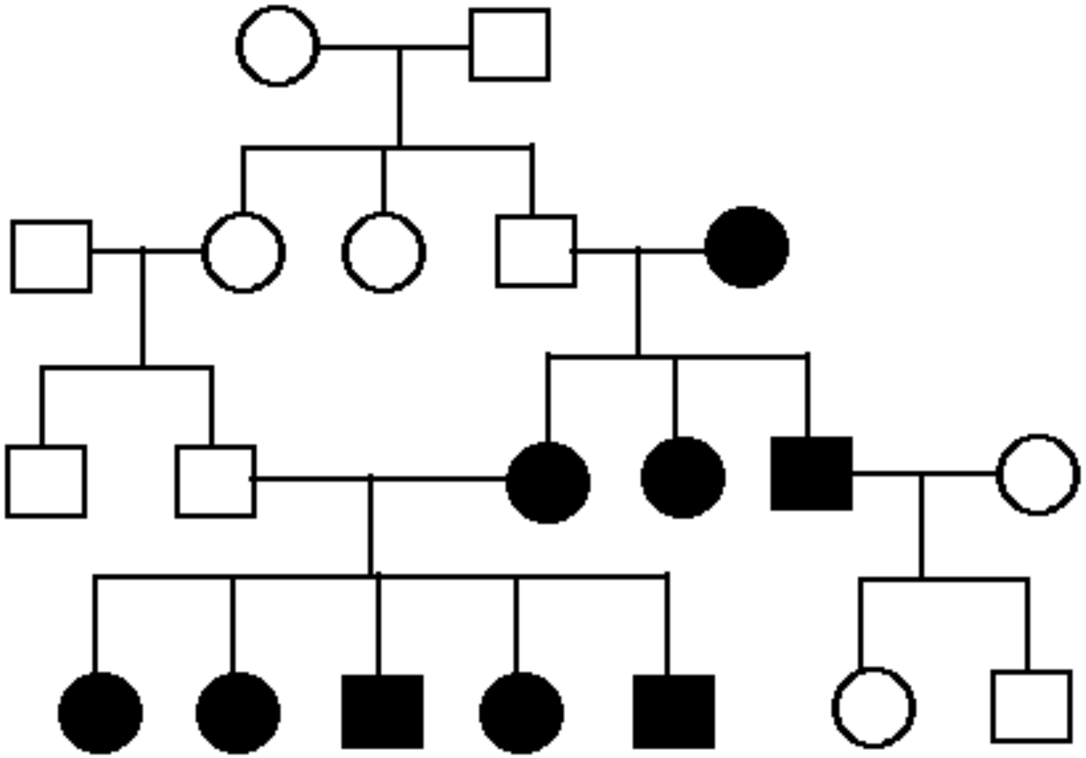
nature methods
ARTICLES
https://doi.org/10.1038/s41592-018-0229-2

Deep generative modeling for single-cell transcriptomics

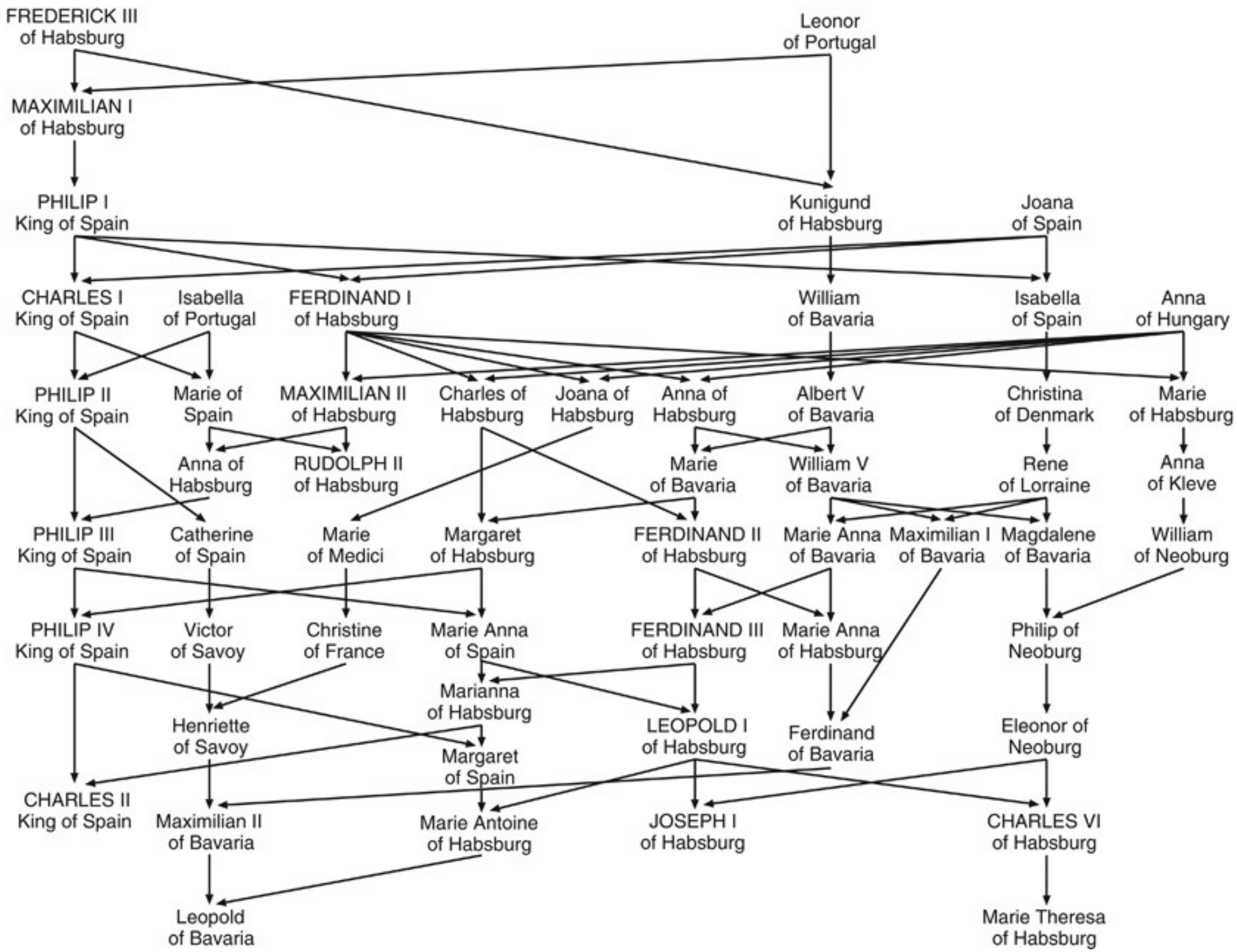
Romain Lopez[1], Jeffrey Regier[1], Michael B. Cole[2], Michael I. Jordan[1,3] and Nir Yosef[1,4,5*]



A PEDIGREE

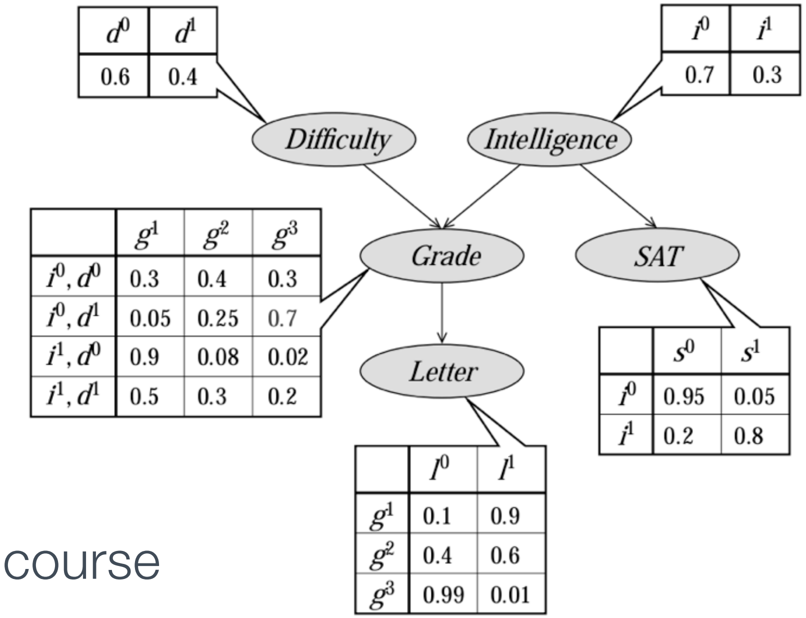


ROYAL DYNASTIES AS HUMAN INBREEDING LABORATORIES: THE HABSBURGS



- Three types of problems
  1. Marginalizing
  2. Learning parameters
  3. Learning structure

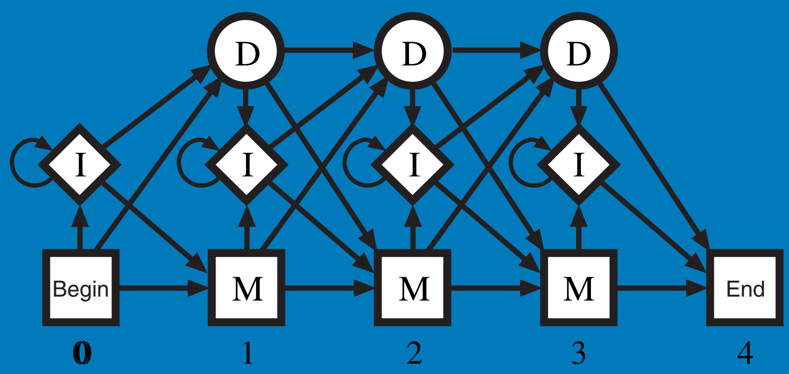
This course



DIRECTED GRAPHICAL MODELS

APPLICATIONS OF HMMS

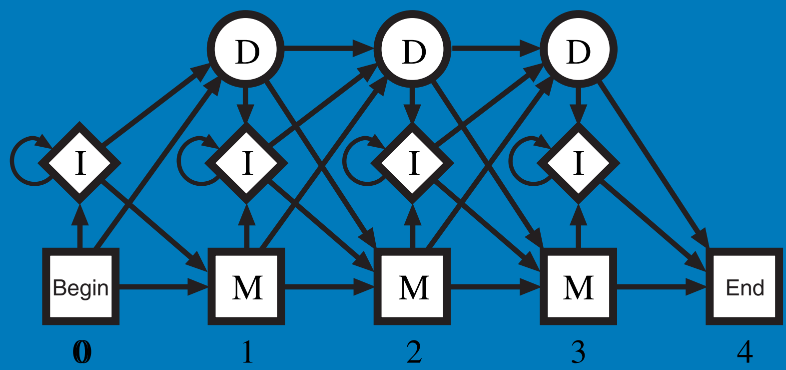
	x	x	.	.	.	x
bat	A	G	-	-	-	C
rat	A	-	A	G	-	C
cat	A	G	-	A	A	-
gnat	-	-	A	A	A	C
goat	A	G	-	-	-	C
	1	2	.	.	.	3



- Automatic speech recognition
- Part of speech tagging
- Gene finding
- Gene family characterization
- Secondary structure prediction

INFERENCE TYPES

	x	x	.	.	.	x
bat	A	G	-	-	-	C
rat	A	-	A	G	-	C
cat	A	G	-	A	A	-
gnat	-	-	A	A	A	C
goat	A	G	-	-	-	C
	1	2	.	.	.	3

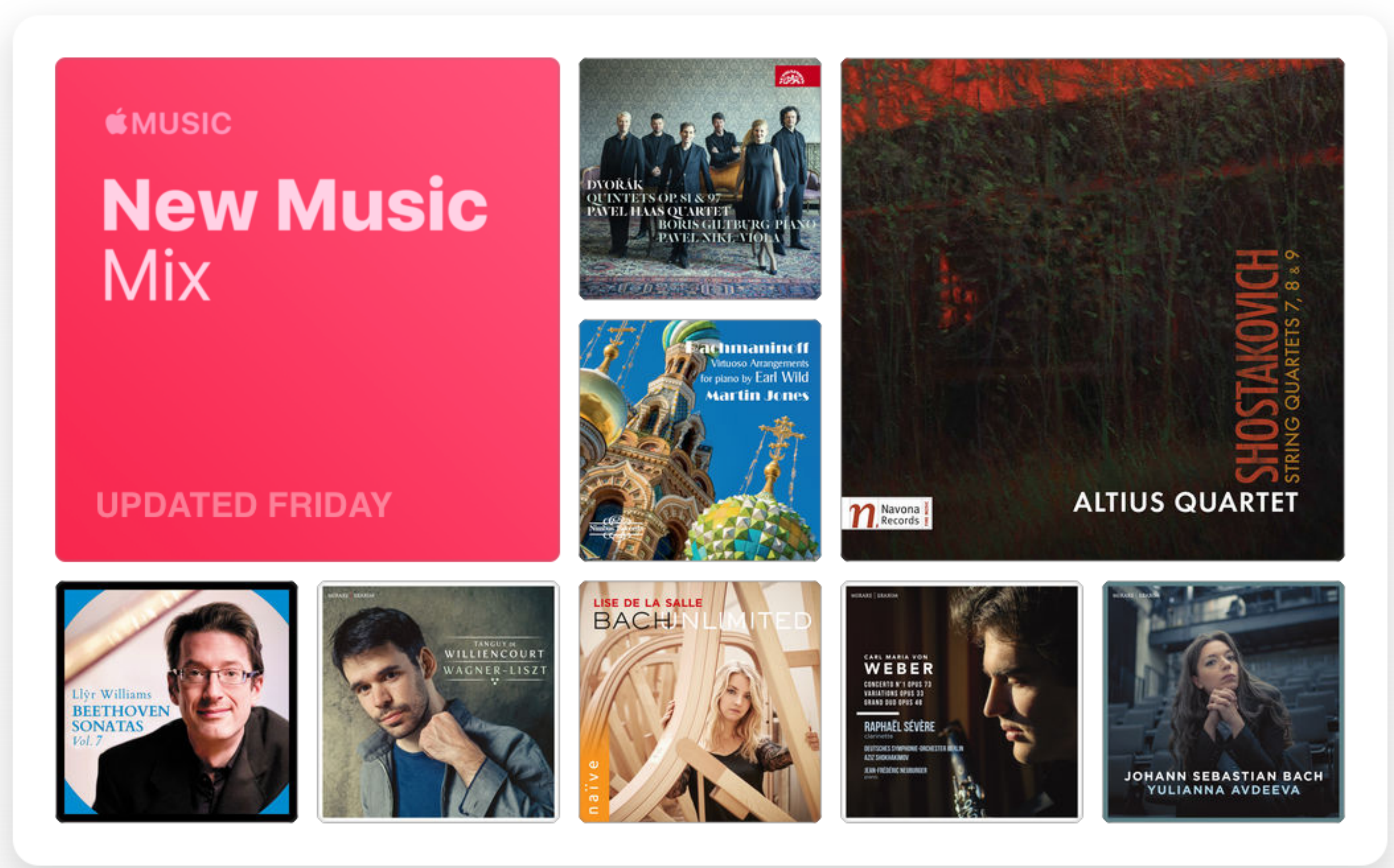


- Probability of data: $p(x_{1:T})$
- Parameters:

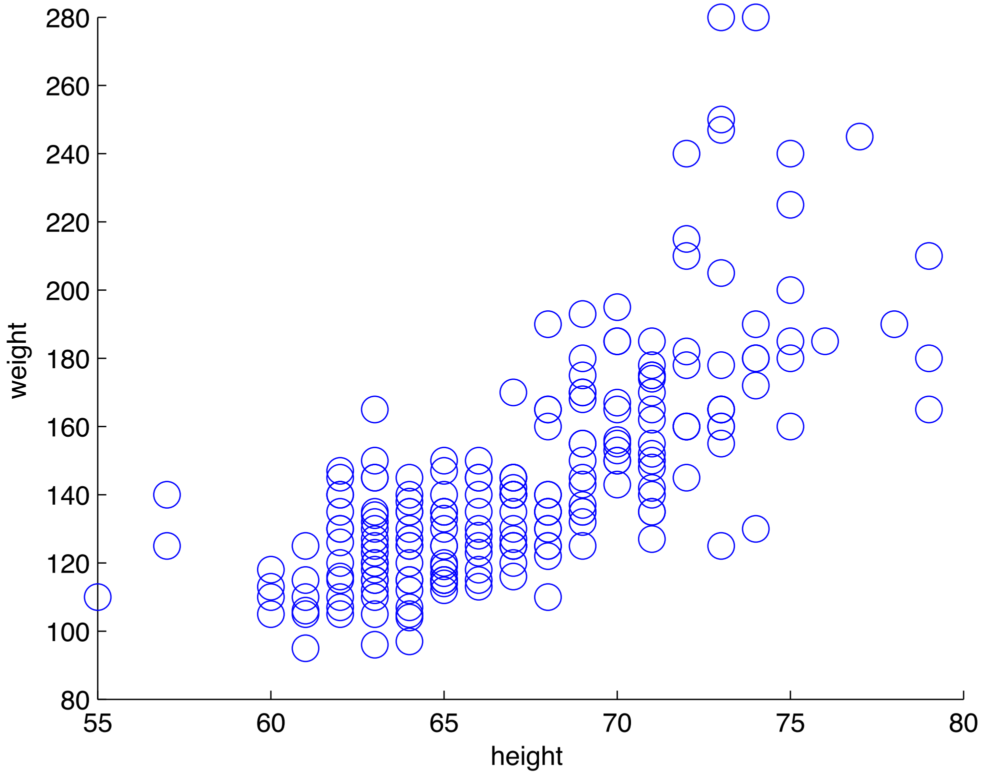
  given D & struct.
- Structure and param.:

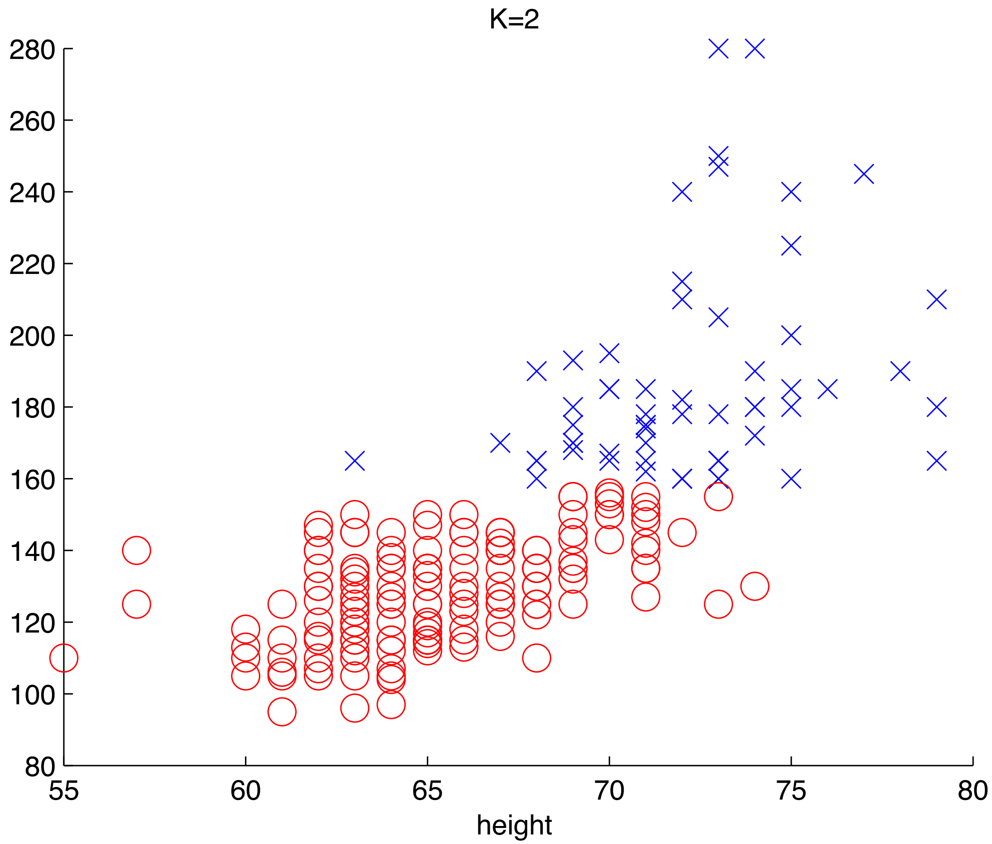
  given D

Using the Expectation Maximisation methodology - EM



COLLABORATIVE FILTERING

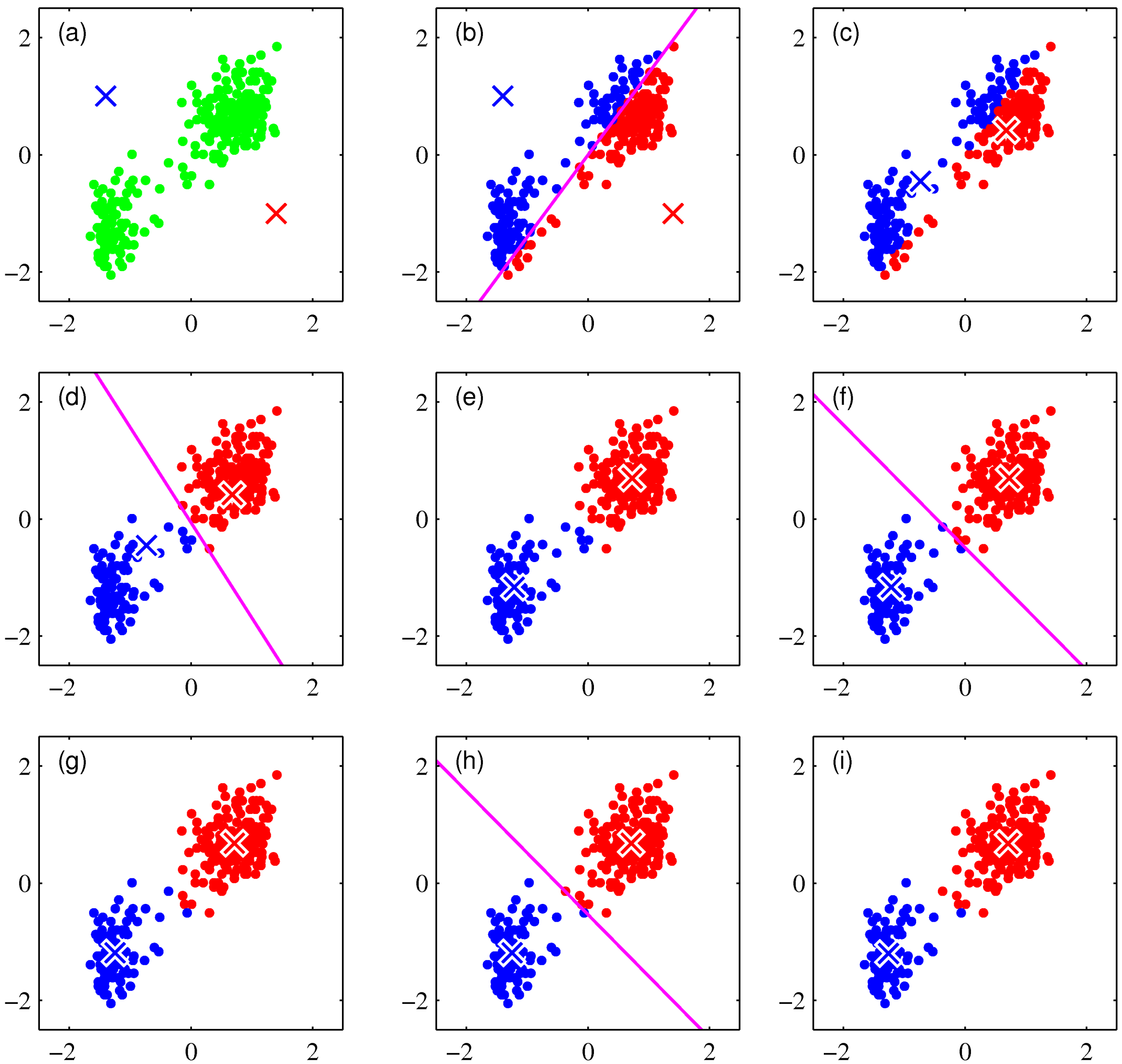




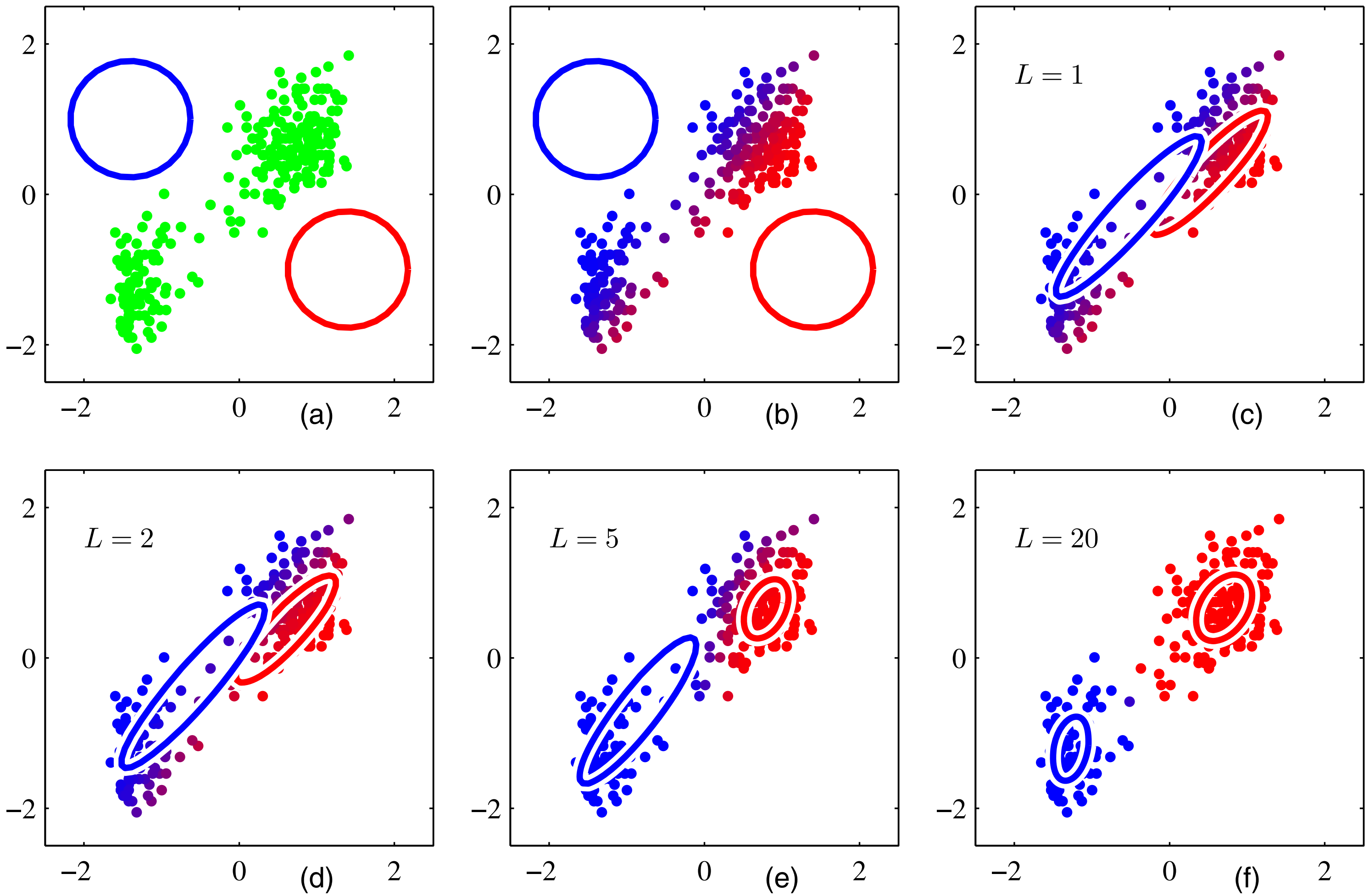
- ★ Each subset should contain similar points
- ★ Pairs of subsets should have dissimilar points.

CLUSTERING

K-MEANS



EM CLUSTERING - CAN BE VIEWED AS SOFT VERSION



VARIATIONAL BAYES - MEAN FIELD

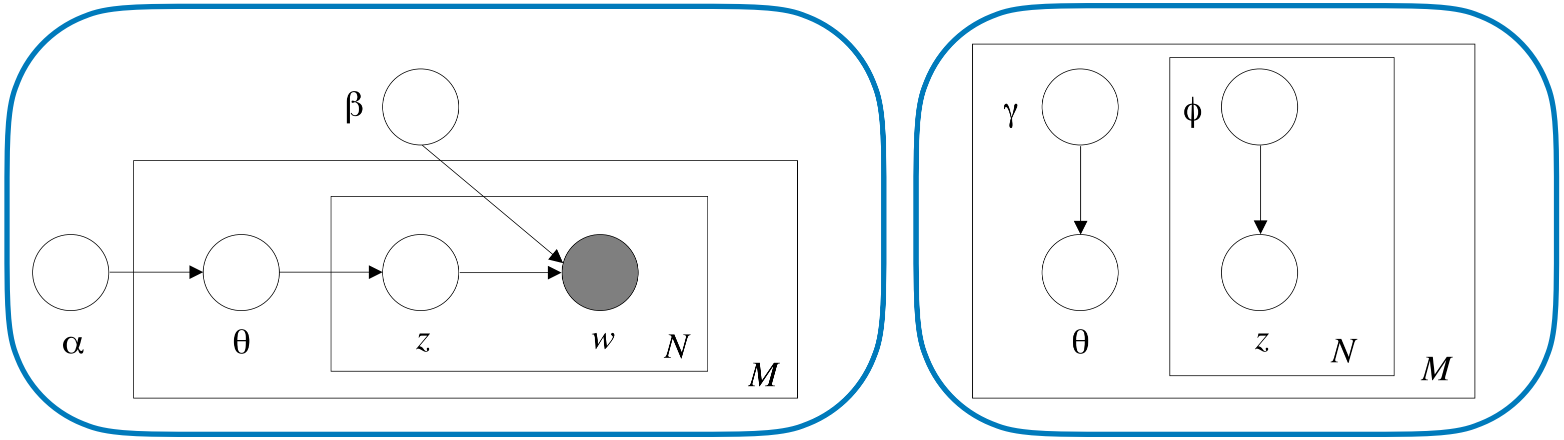


Figure 5: (Left) Graphical model representation of LDA. (Right) Graphical model representation of the variational distribution used to approximate the posterior in LDA.

- Variational Bayes (VB)
  - Technique for approximating a posterior

KL

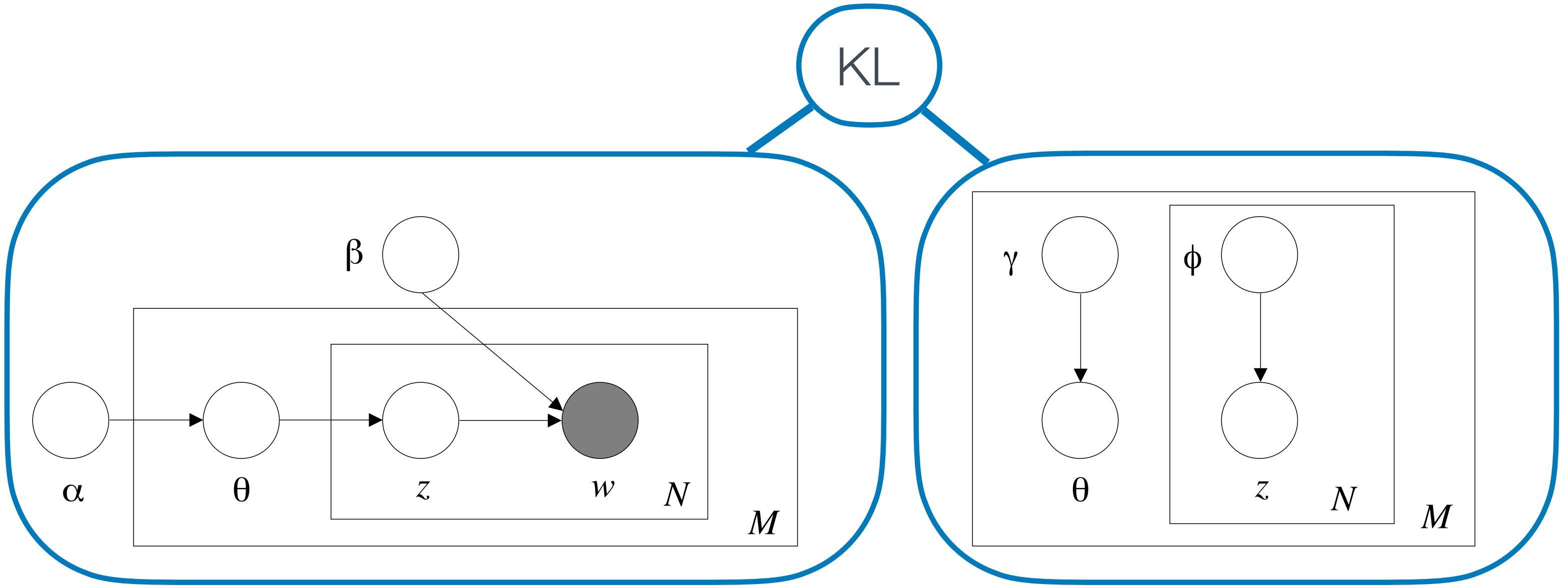


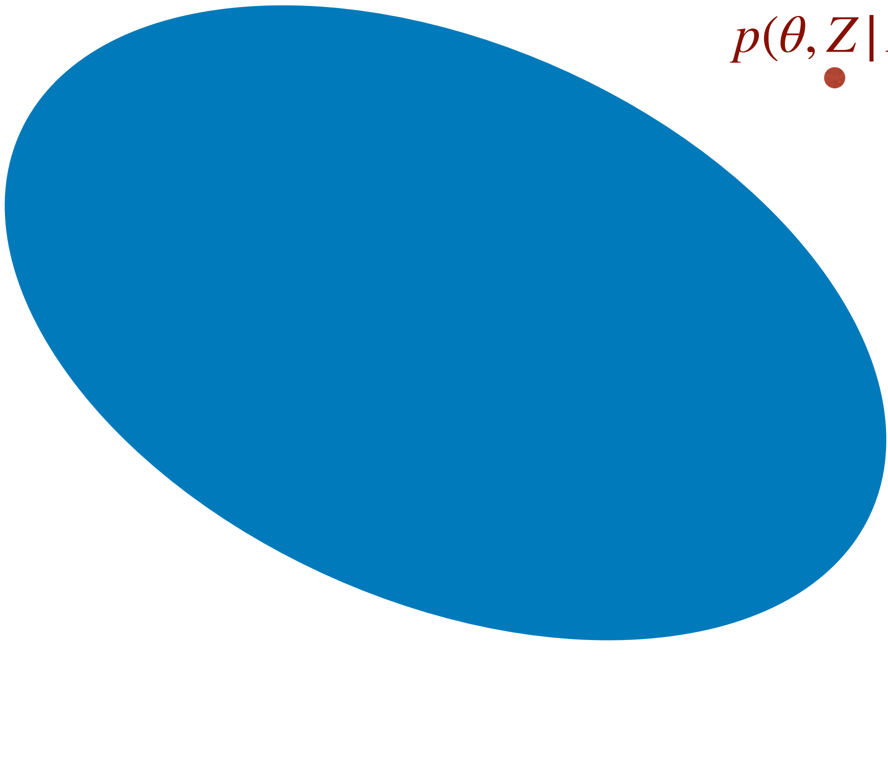
Figure 5: (Left) Graphical model representation of LDA. (Right) Graphical model representation of the variational distribution used to approximate the posterior in LDA.

- Variational Bayes (VB)
  - Technique for approximating a posterior

MINIMIZING KL-MAXIMIZING ELBO

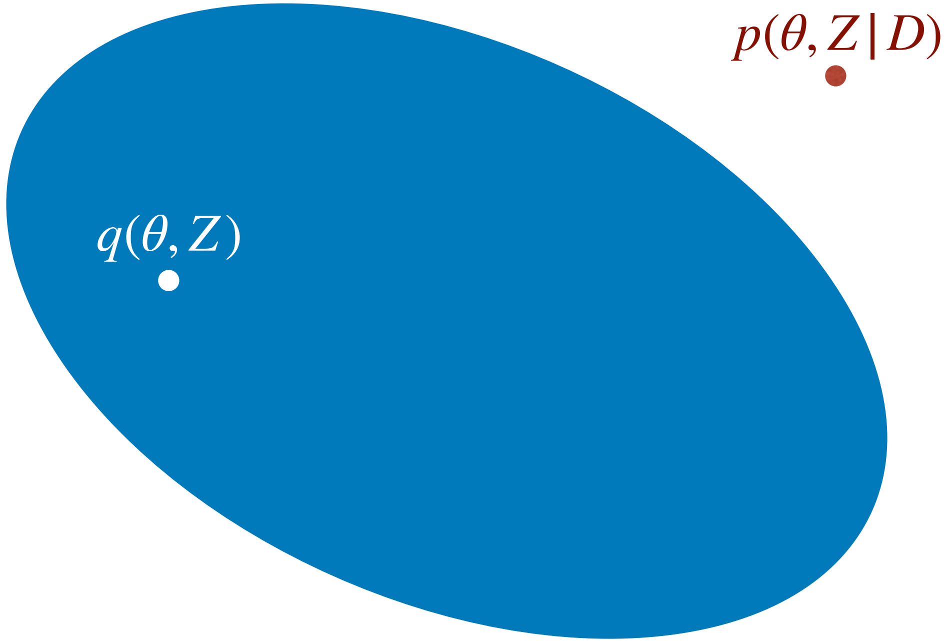
$q(\theta, Z) = q_\theta(\theta) q_Z(,Z)$

$p(\theta, Z \mid D)$

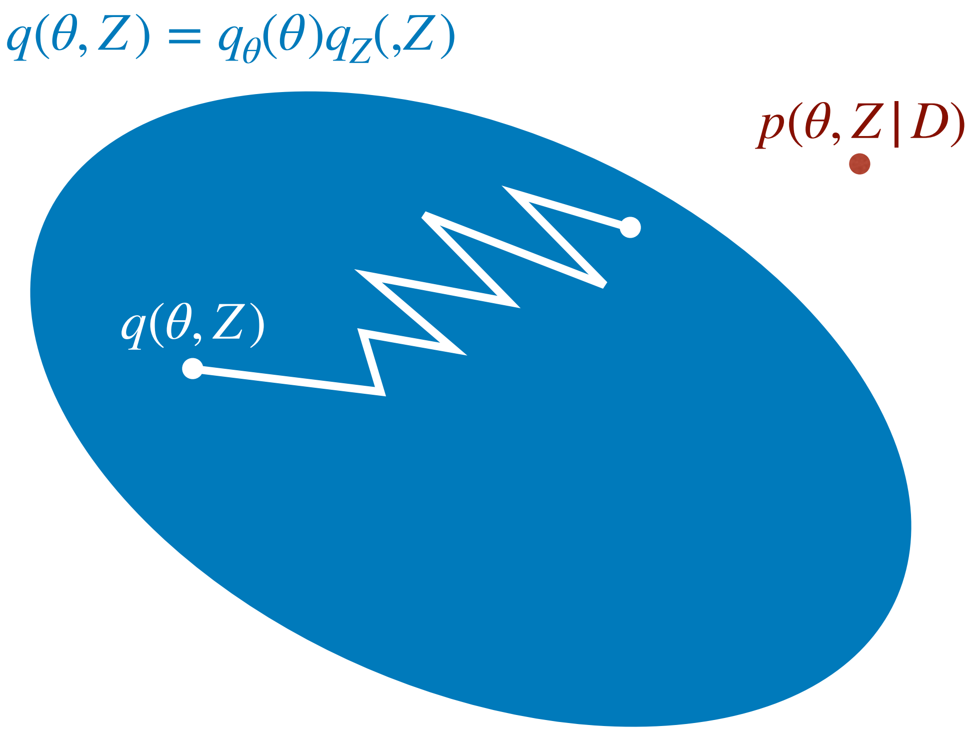


MINIMIZING KL-MAXIMIZING ELBO

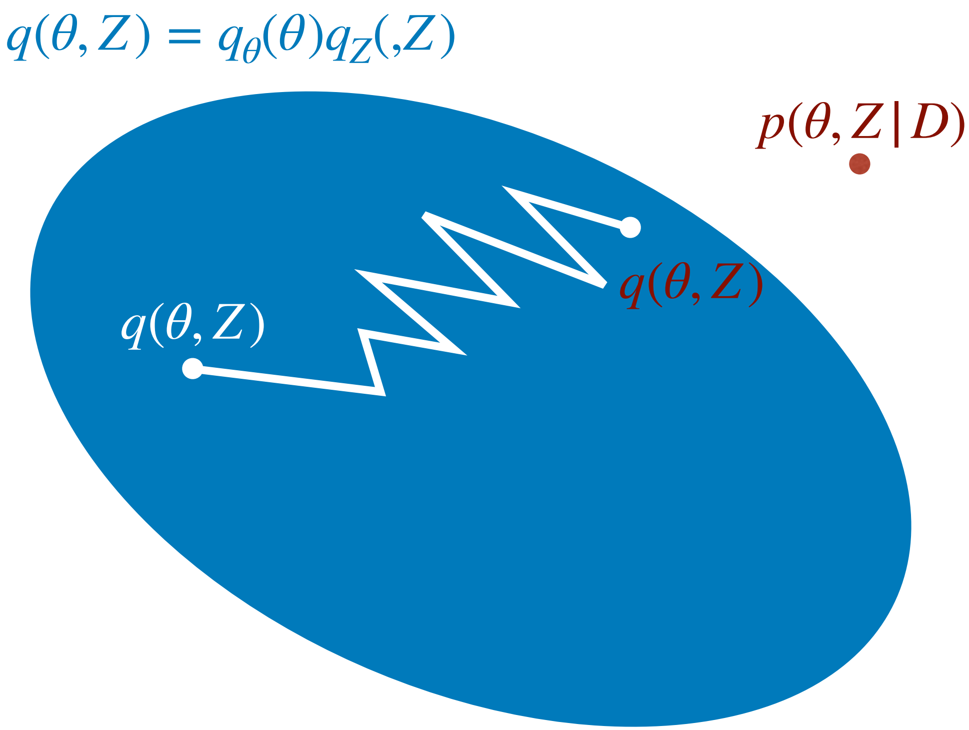
$q(\theta, Z) = q_\theta(\theta) q_Z(, Z)$



MINIMIZING KL-MAXIMIZING ELBO



MINIMIZING KL-MAXIMIZING ELBO



ELBO AGAIN - VARIATIONAL AUTOENCODER

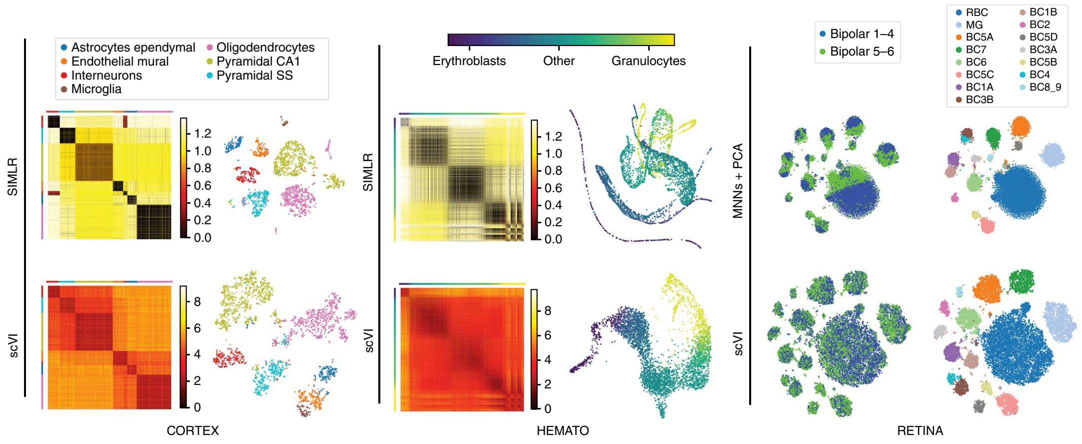
Deep generative modeling for single-cell transcriptomics

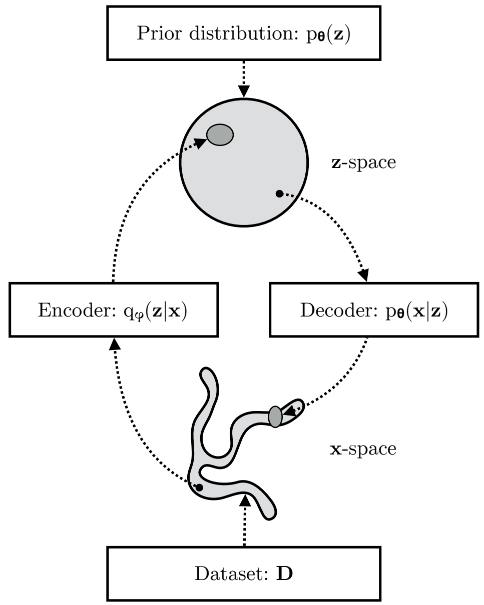
Romain Lopez[1], Jeffrey Regier[1], Michael B. Cole[2], Michael I. Jordan[1,3] and Nir Yosef[1,4,5*]

Auto-Encoding Variational Bayes

Diederik P. Kingma
Machine Learning Group
Universiteit van Amsterdam
dpkingma@gmail.com

Max Welling
Machine Learning Group
Universiteit van Amsterdam
welling.max@gmail.com





THE END