Advanced Topics in Machine Learning

Machine Learning and Computational Statistics (DSC6135)

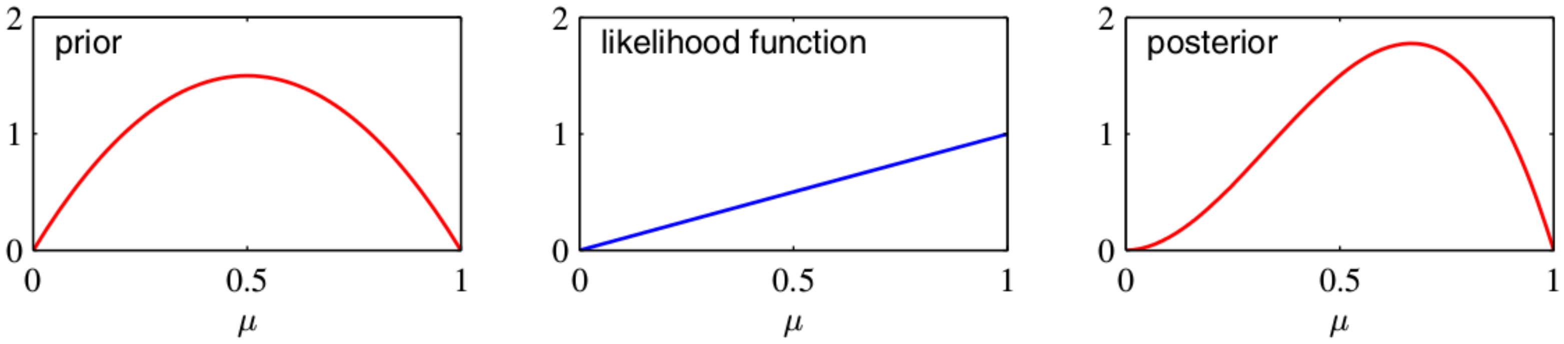
Instructors: Weiwei Pan (Harvard), Javier Zazo (Harvard), Melanie F. Pradier (Harvard)

Bayesian Modeling

- Probability = degree of belief (in contrast with frequentist definition)
- **Bayes Rule** (combines prior knowledge with data evidence)

$$p(W|X) = \frac{p(X|W)p(W)}{p(X)}$$

$$posterior = \frac{likelihood \cdot prior}{evidence}$$



Note: evidence also called marginal likelihood.

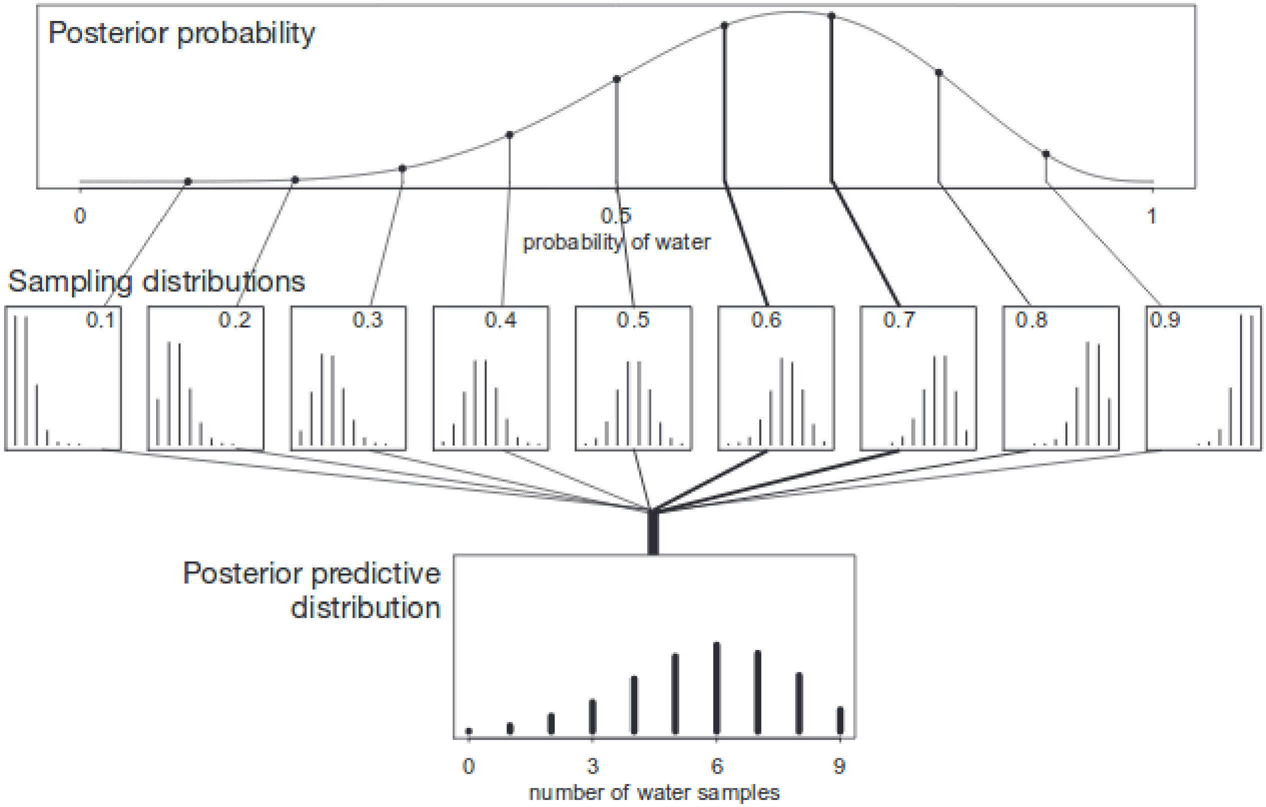
Estimators for model parameters:

- Maximum Likelihood Estimator
$$W_{ML} = \underset{W}{\operatorname{argmax}}\ p(X|W)$$

- Maximum A Posteriori
$$W_{MAP} = \underset{W}{\operatorname{argmax}}\ p(W|X)$$

Beyond point estimates for W:

Posterior predictive:



What we have learned...

- **Key question**: Given some data, how to choose the best model given the data?

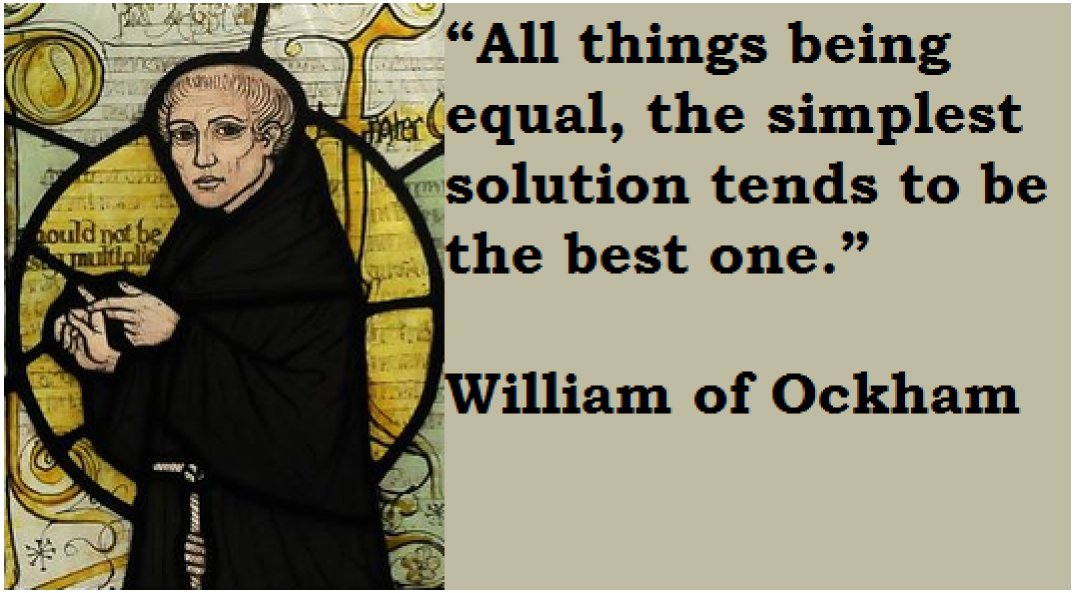
- Simplest approach: find best model by minimizing error (e.g., MSE)

  - If function is too simple, bias will be large (underfitting) -> let's fit a flexible model!

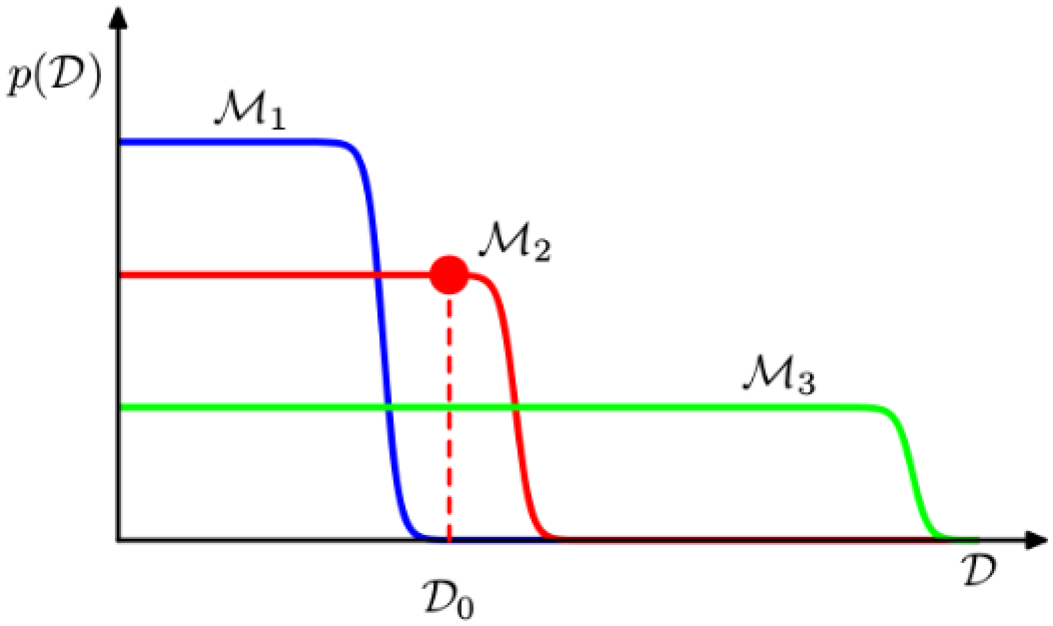
  - If function is too complex, variance will be large (overfitting) -> let's decrease variance!

    1) Bagging
    2) Regularization
    3) Prior Knowledge

Model Selection: Occam's Razor [Bishop Book]



Model Selection: Occam's Razor [Bishop Book]



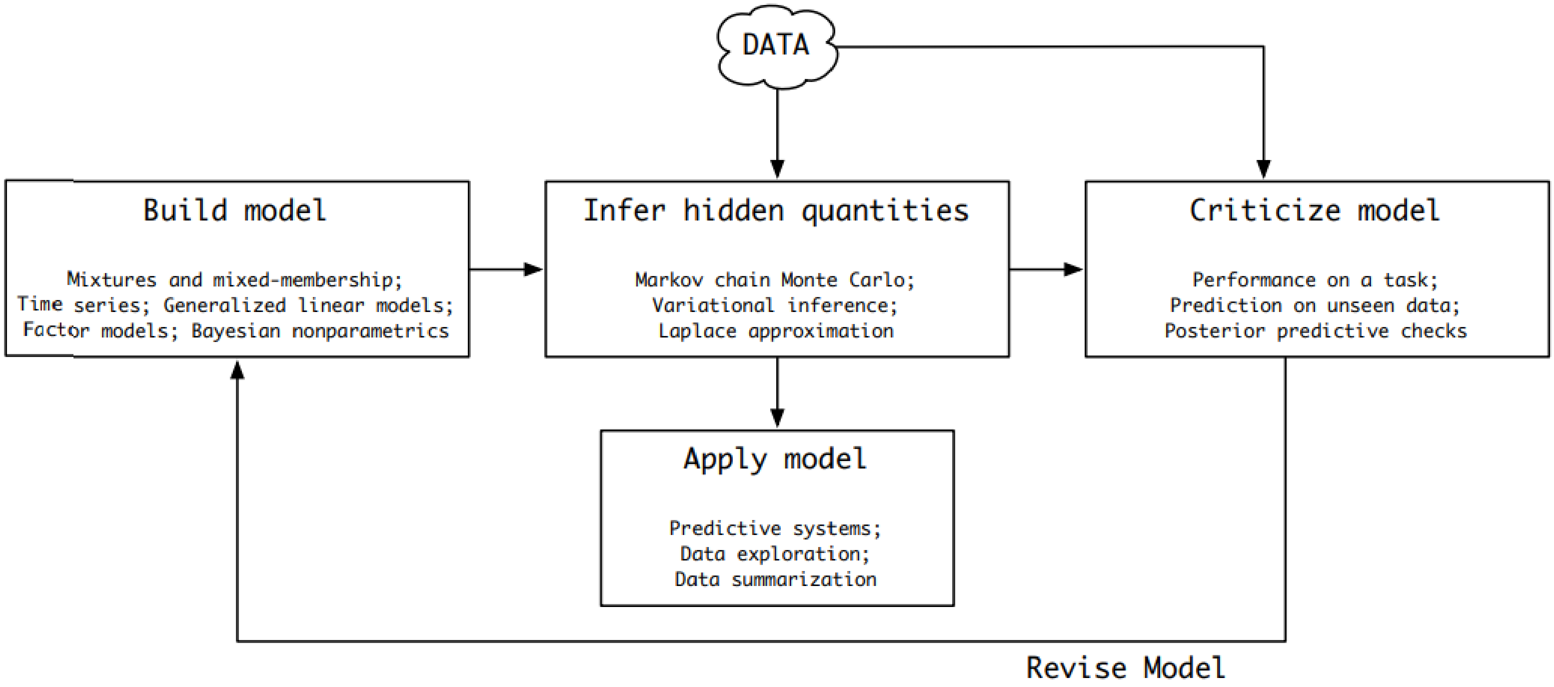
What we have learned...

- **Supervised learning**
  - regression (e.g., kNN, linear, polynomial regression, neural networks)
  - classification (e.g., logistic regression, polynomial logistic, decision trees, random forest, neural networks)

- **Unsupervised learning**

  - clustering (K-means, Gaussian Mixture Models)

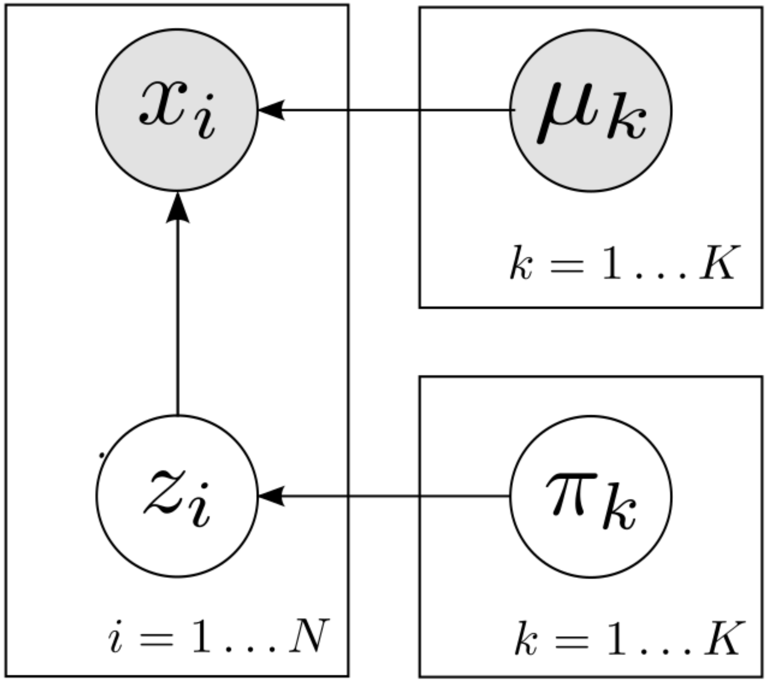
Machine Learning Pipeline



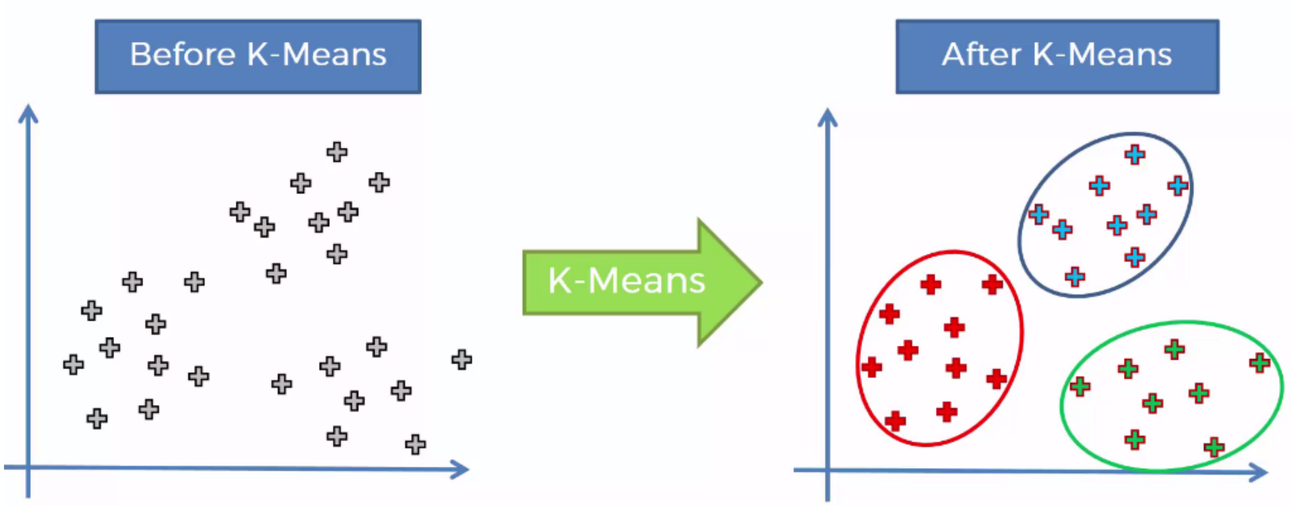
What Next?

Generative Models

- Graphical representation of random variables and their dependencies.
- Example: Mixture Models (see blackboard)



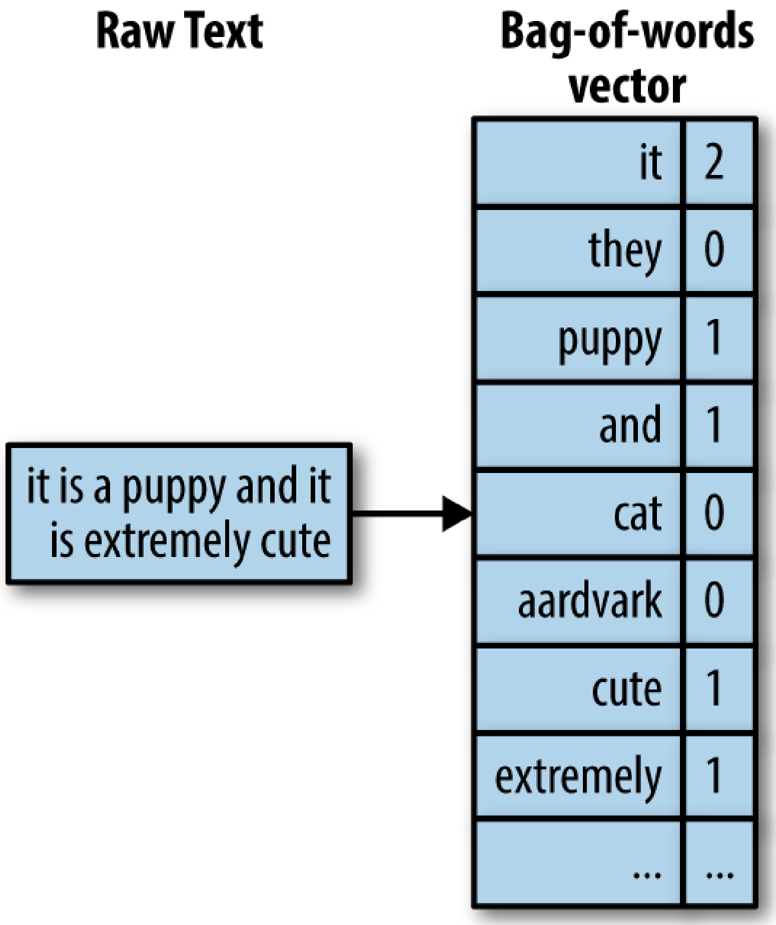
Clustering Documents



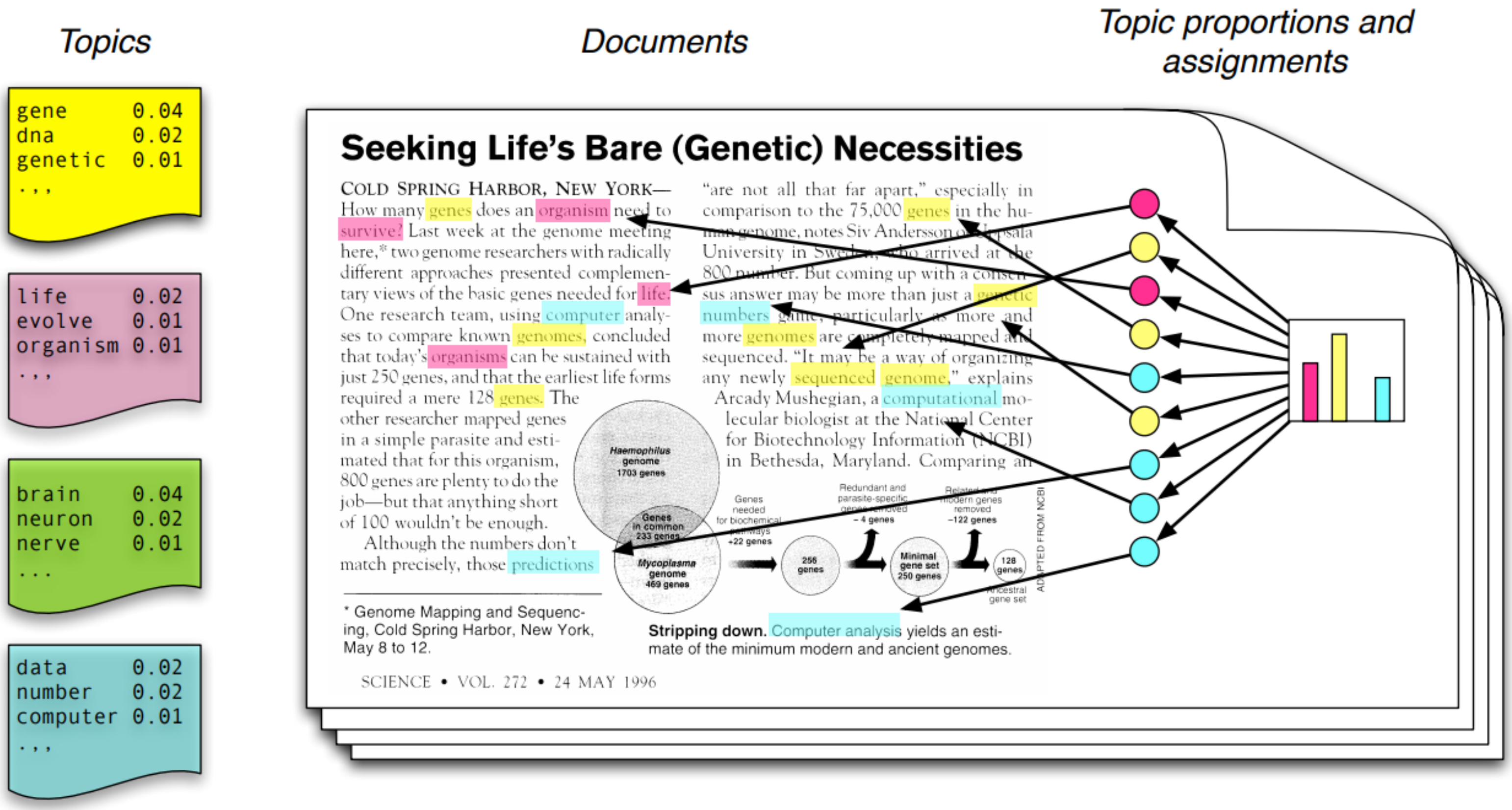
Which challenges do you foresee?

- Stop words: "THE lady was happy BECAUSE SHE saw A rainbow"
  - One solution: remove "non-informative words" (TF-IDF: term frequency inverse document frequency)
- One single category per document is assumed!

Representation of text data: ”bag-of-words”

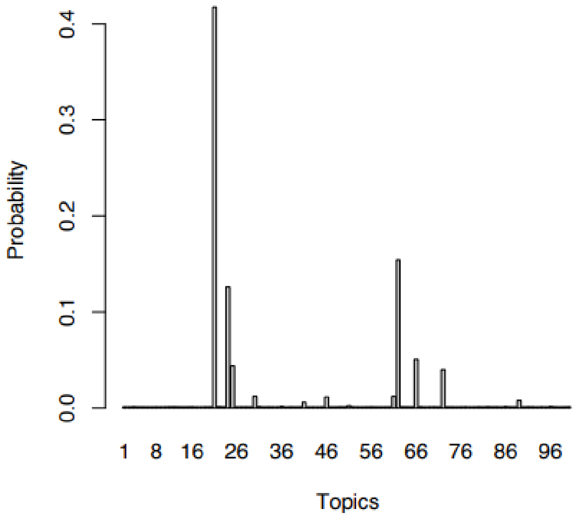


Assumptions behind Topic Models



Blei, David M.. "Introduction to Probabilistic Topic Models." (2010).

Topic Models

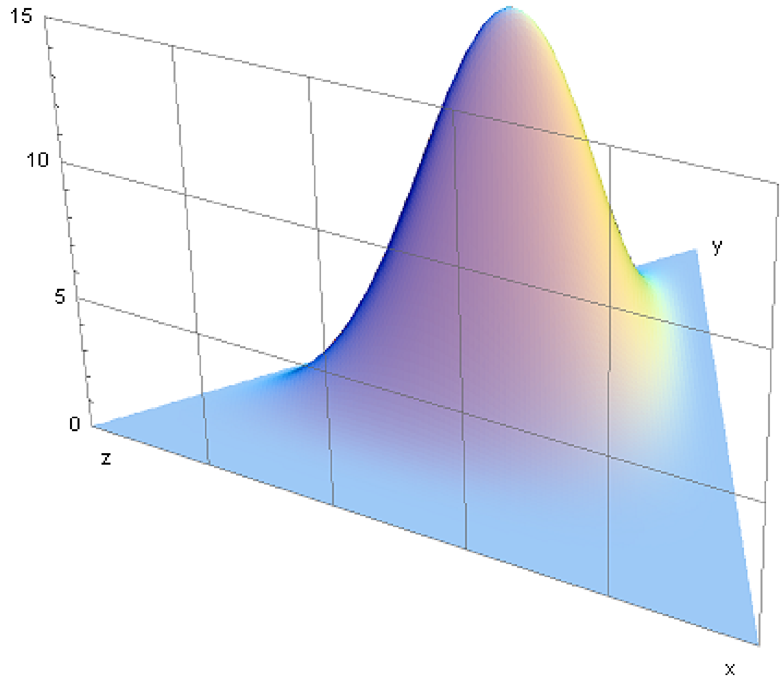


"Genetics"	"Evolution"	"Disease"	"Computers"
human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

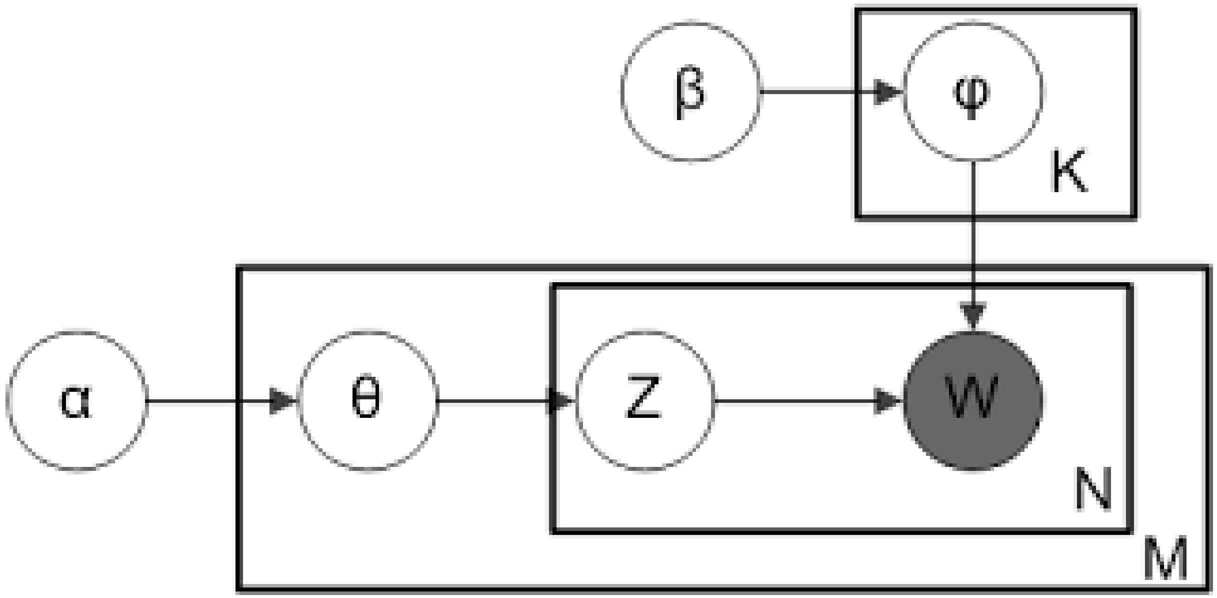
Blei, David M.. "Introduction to Probabilistic Topic Models." (2010).

Let's formalize this model!!

Dirichlet Distribution

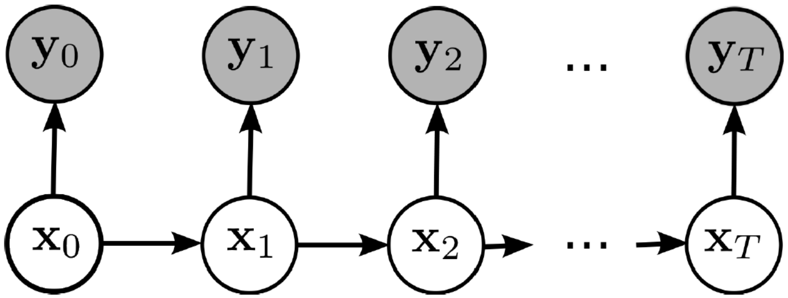


Topic Models



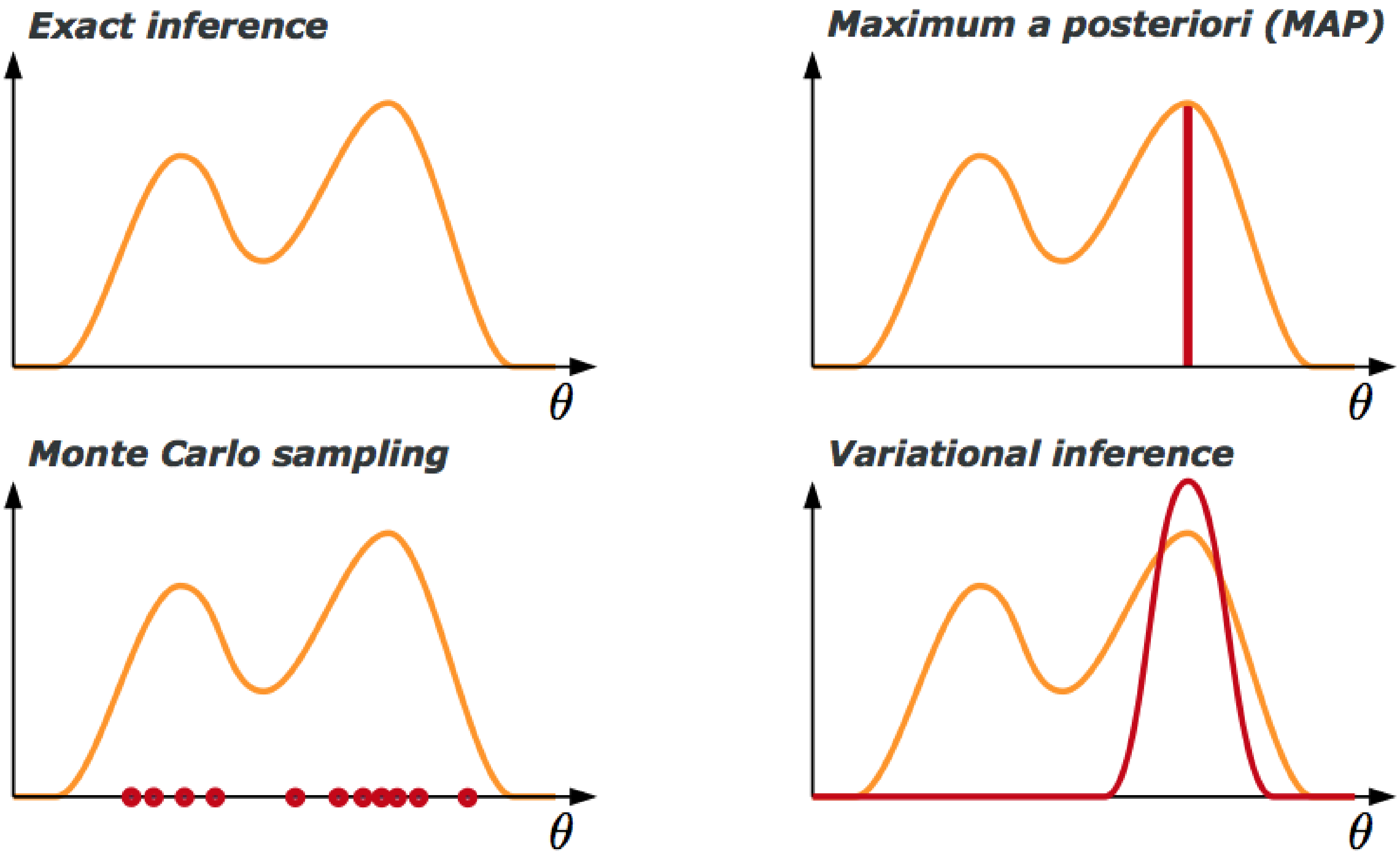
Hidden Markov Models

Hidden Markov Models



- Here: y_i: observed, x_i: hidden
- Markov assumptions:
  - current observation only depends on current hidden state
  $$p(Y_t|X_{0:(t-1)}) = p(Y_t|X_t)$$
  - current hidden state only depends on past hidden state
  $$p(X_t|X_{0:(t-1)}) = p(X_t|X_{(t-1)})$$

Approximate Bayesian Inference



Summary

Probabilistic modeling

- makes model assumptions **explicit**
- encode our "beliefs" on the data
- Examples: topic models, hidden markov models, etc...

```
In [ ]:
```