The Gaussian Mixture and Expectation-Maximization

Alexander G. Ororbia II

Introduction to Machine Learning

CSCI-635

11/1/2023

Generation vs. Discrimination

Generative Models

- Represent both the data and the labels
- Often, makes use of conditional independence and priors
- Examples
  - Naïve Bayes classifier
  - Bayesian network
  - **Gaussian mixtures**
  - Single/Dual-wing harmonium
  - Variational autoencoder
- Models of data may apply to future prediction problems

Discriminative Models

- Learn to directly predict the labels from the data
- Often, assume a simple boundary (e.g., linear)
- Examples
  - Logistic regression
  - Multinoulli regression
  - SVM
  - Decision tree / An ensemble
  - MLP
- Often easier to predict a label from the data than to model the data

White board time!

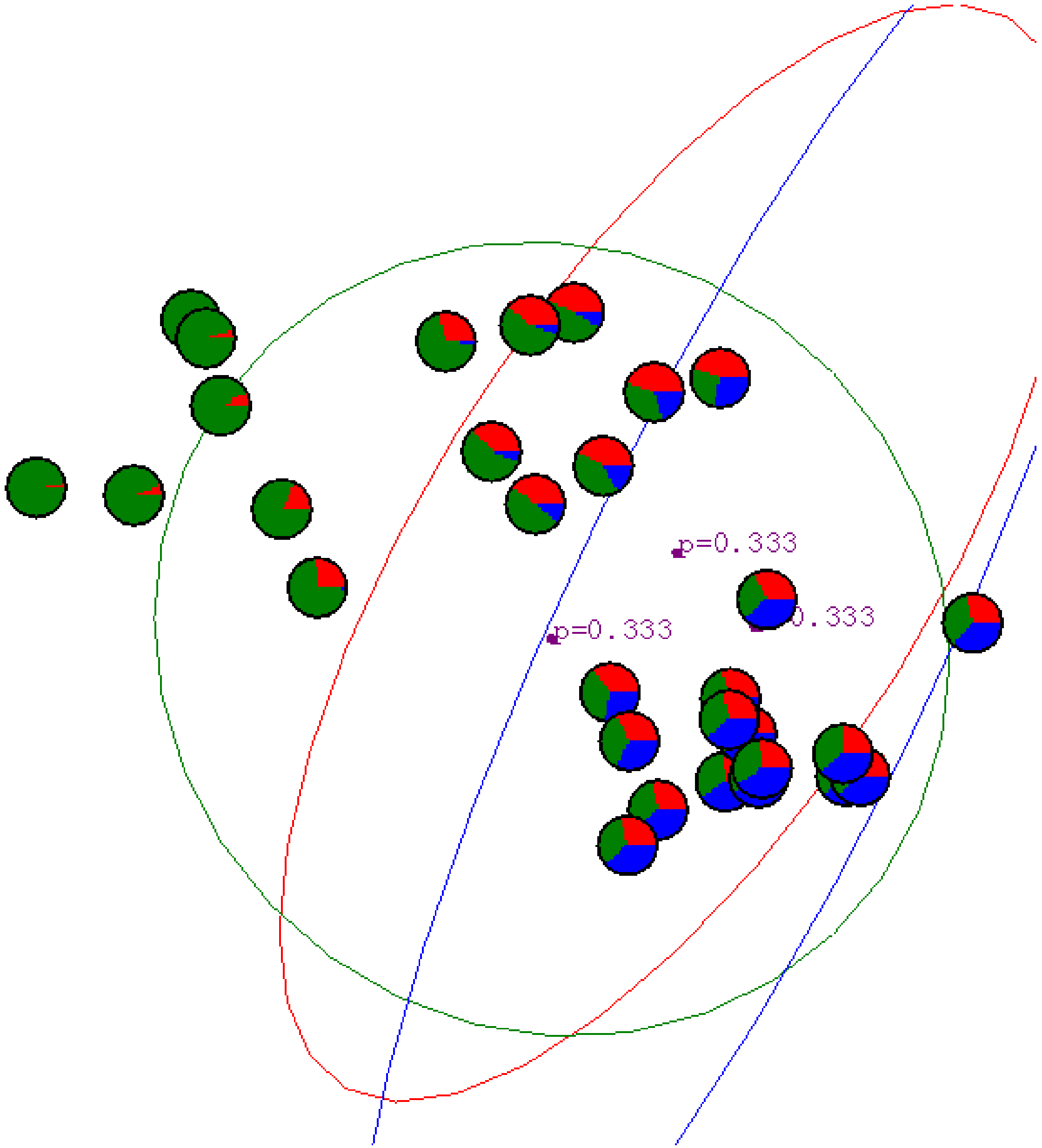
Deriving/crafting a GMM and its E-M fitting process

Expectation Maximization (EM)

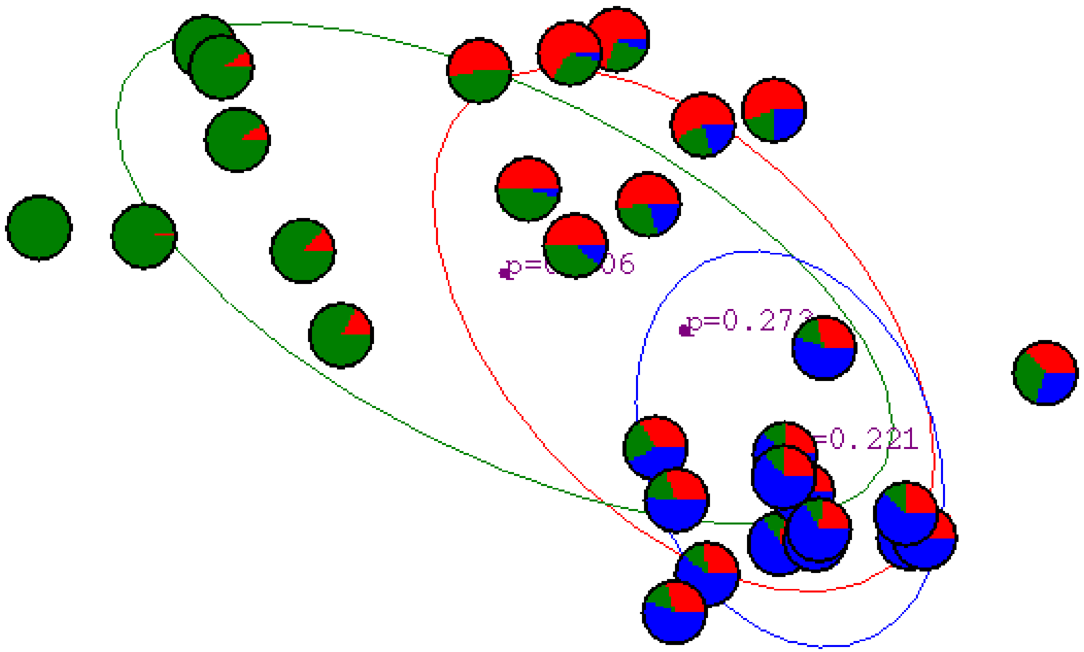
- Training of GMMs with latent variables accomplished via Expectation Maximization
  - **Step 1:** Expectation (**E**-step)
    - Evaluate the “responsibilities” of each cluster with the current parameters
  - **Step 2:** Maximization (**M**-step)
    - Re-estimate parameters using the existing “responsibilities”

- Similar to k-means (clustering) training (Friday)

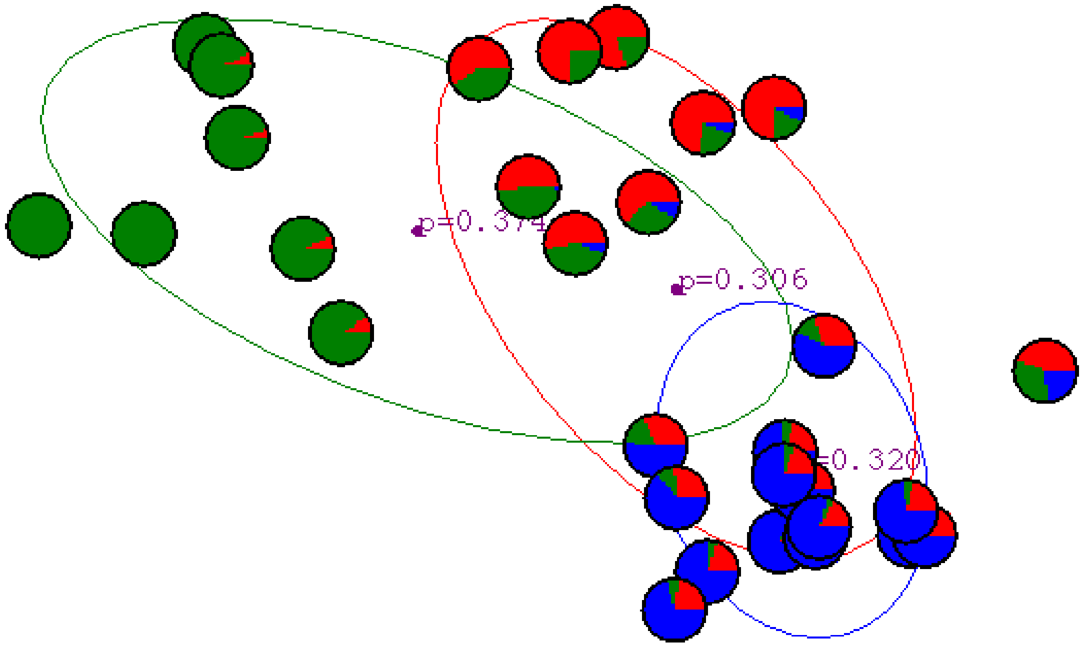
Gaussian Mixture Example: Start



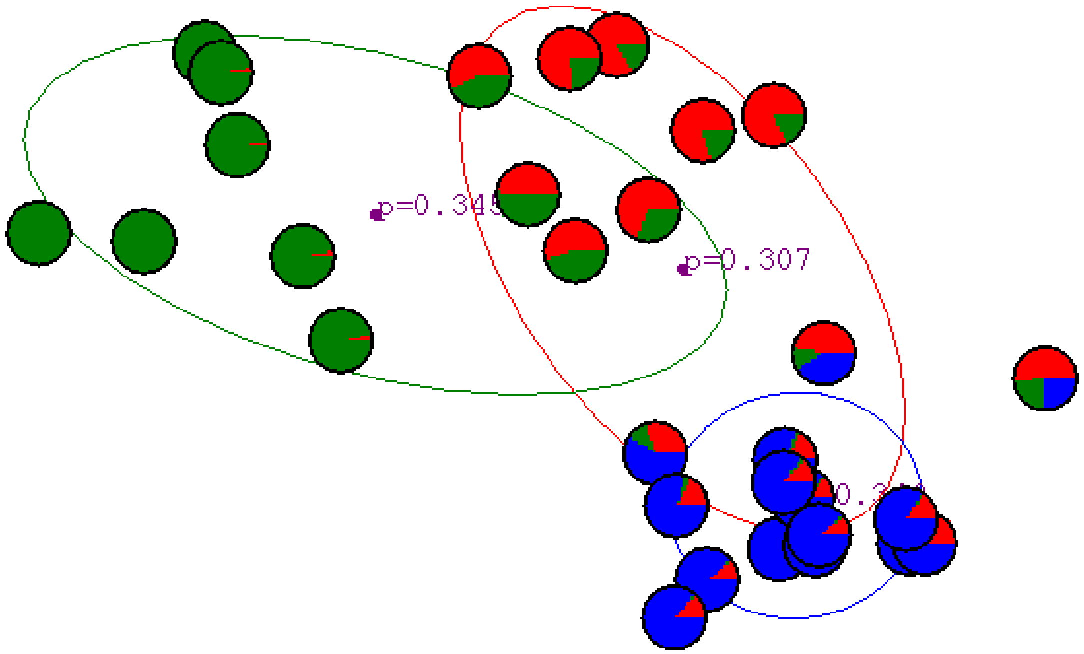
After first iteration



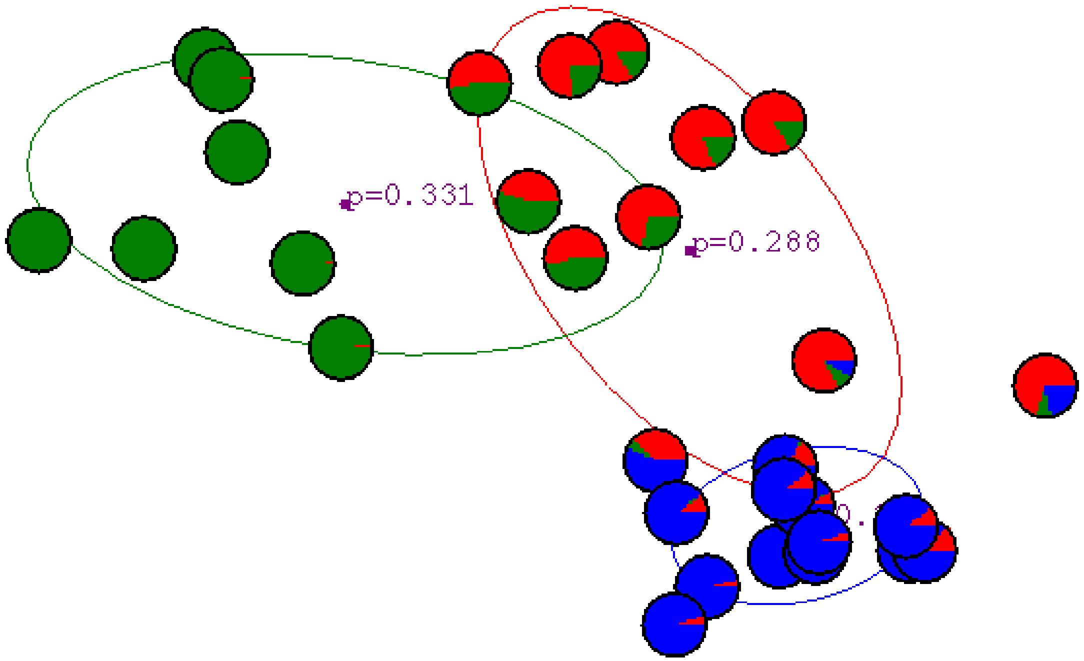
After 2nd iteration



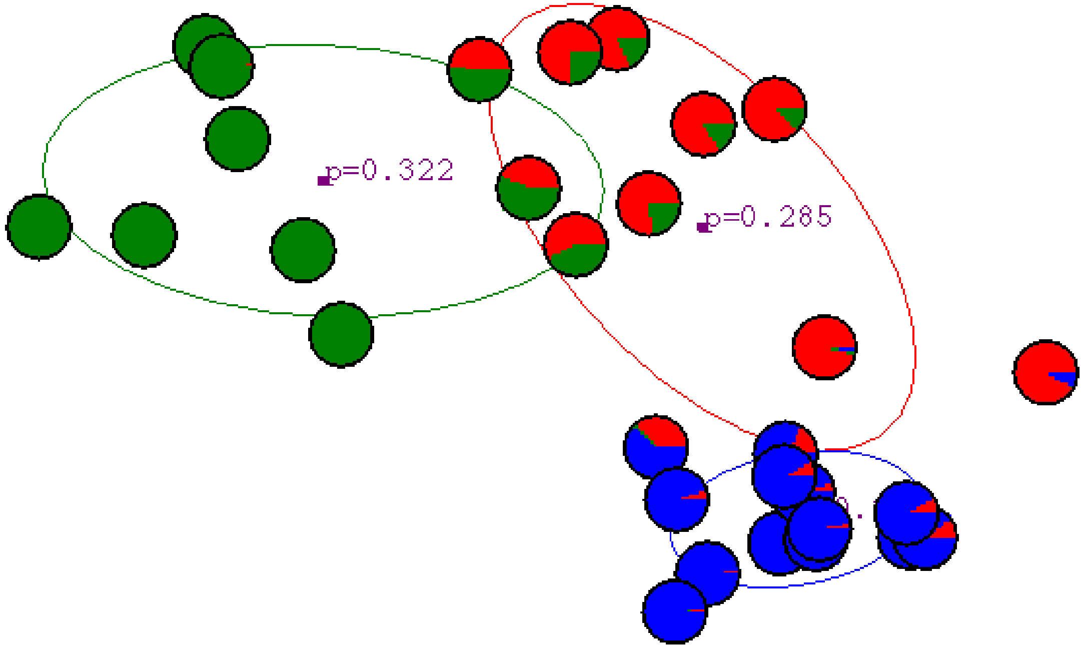
After 3rd iteration



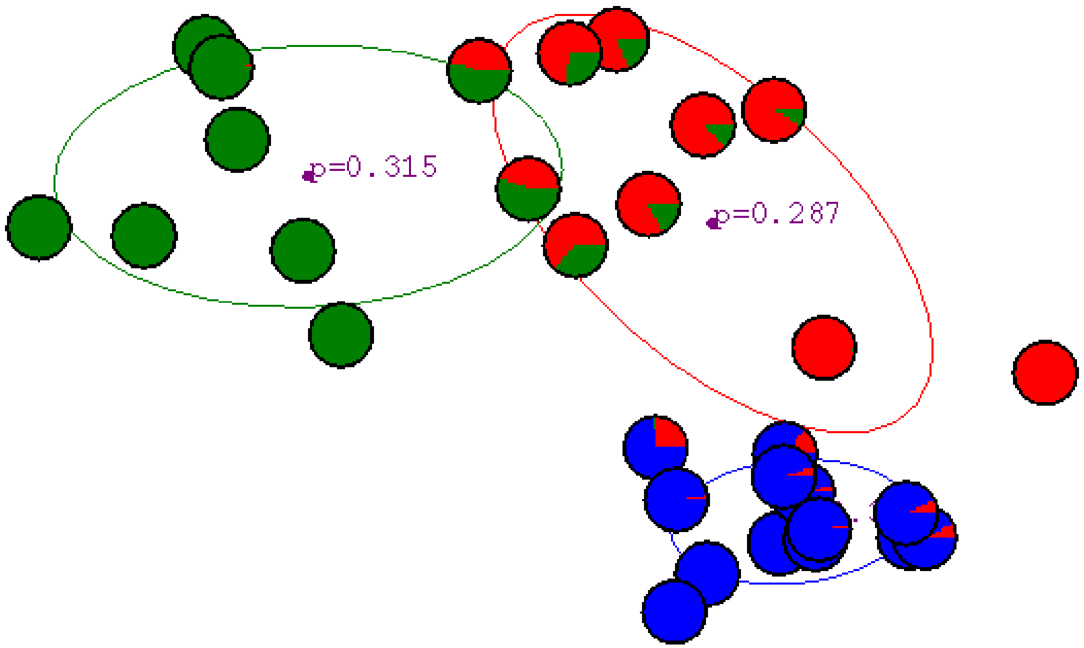
After 4th iteration



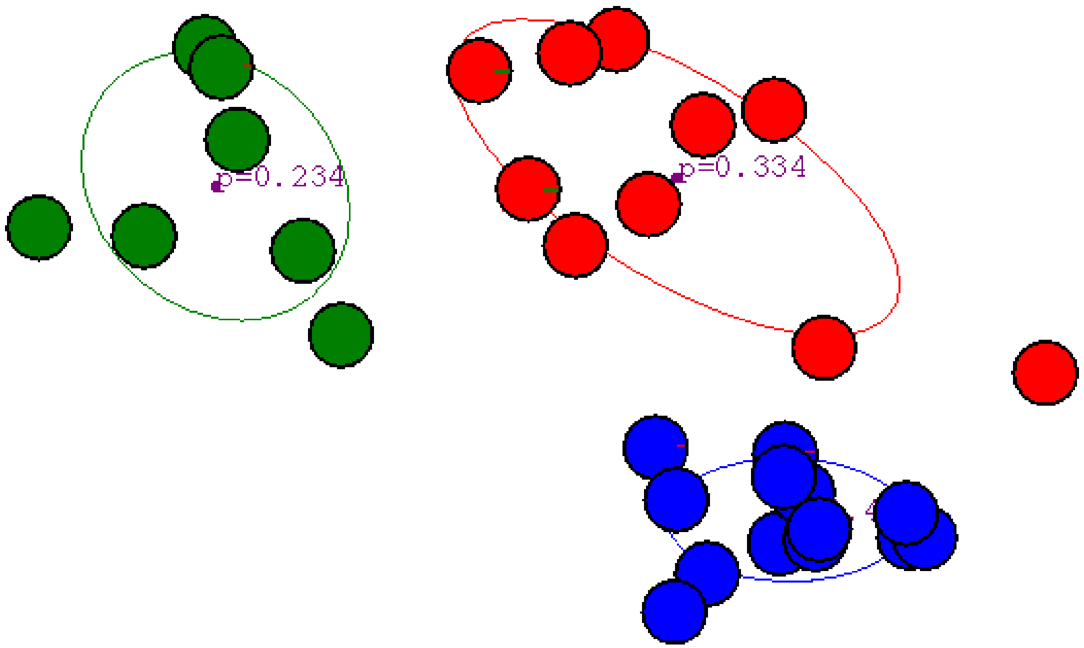
After 5th iteration



After 6th iteration



After 20th iteration



Mixture Distributional Models

- You can construct any mixture model so long as you:
  – Can write down the PMF/PDF of your distribution
  – Can derive/produce MLE estimates of distributional parameters
- Example: Exponential Mixture Model (EMM)

Likelihood function:

$$\mathcal{L}(X;\theta) = \Pi_{i=1}^{N} \lambda \exp(-\lambda x_i) = \lambda^N \exp\Big(-\lambda N \frac{1}{N} \sum_i x_i\Big) = \lambda^N \exp\Big(-\lambda \sum_i x_i\Big)$$

MLE Estimator:

$$\hat{\lambda} = \frac{1}{(1/N)\sum_i \mathbf{x}_i} = \frac{N}{\sum_i \mathbf{x}_i}$$

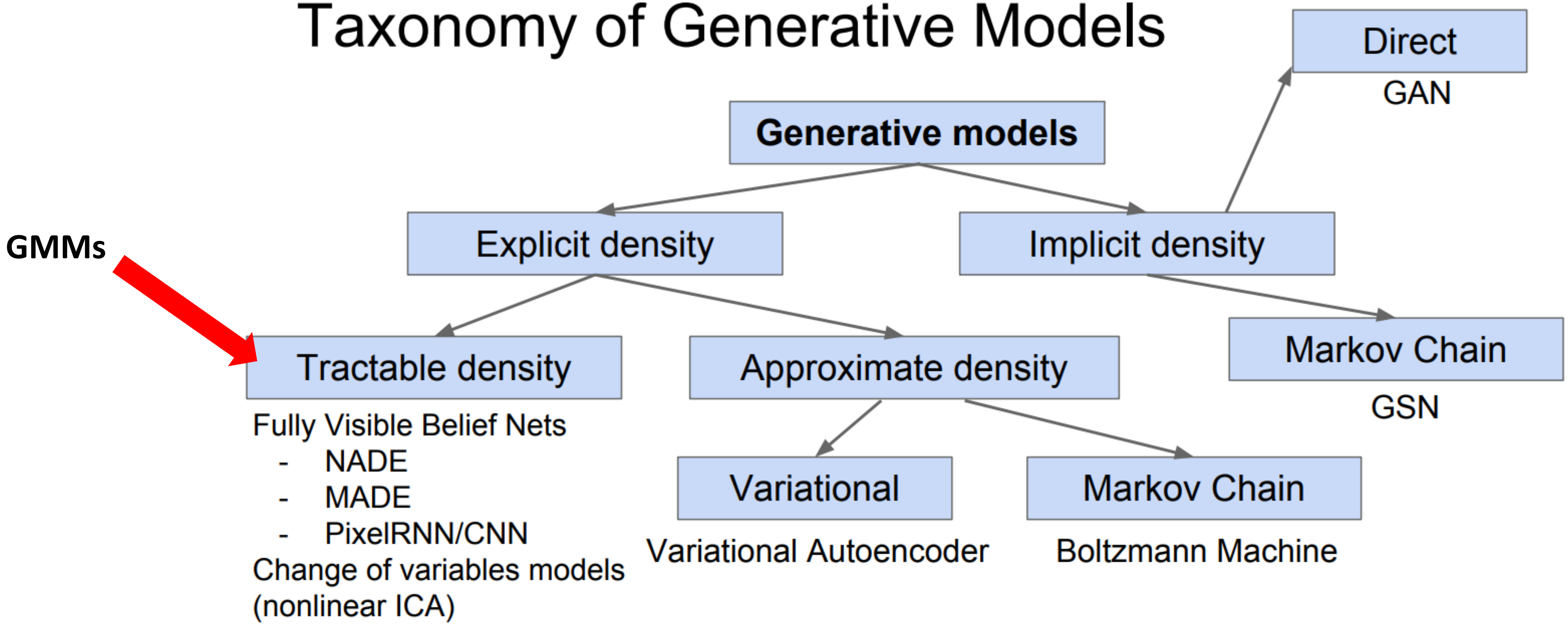
Mixture Distributional Models

- Other mixture models
  - Bernoulli mixture model (BMM)
  - Categorical mixture model (CMM) – discrete variables w/ dict size V
  - Etc., etc.
- Bayesian forms of any mixture model
  - Impose priors over distributional parameters and mixture weights (priors should be conjugate – prior & posterior in same family, e.g., exponential family)

GMM Applications

- Feature extraction from speech data → speech recognition systems
- Used extensively in object tracking of multiple objects
  - Where # of mixture components & their means predict object locations at each frame in video sequence
  - EM algorithm used to update component means over time as video frames update -> allows object tracking
- And much more!
  - Applications similar to clustering too!

Taxonomy of Generative Models



QUESTIONS?

QUESTIONS?