E9 205 Machine Learning for Signal Processing

Summary of Generative Models

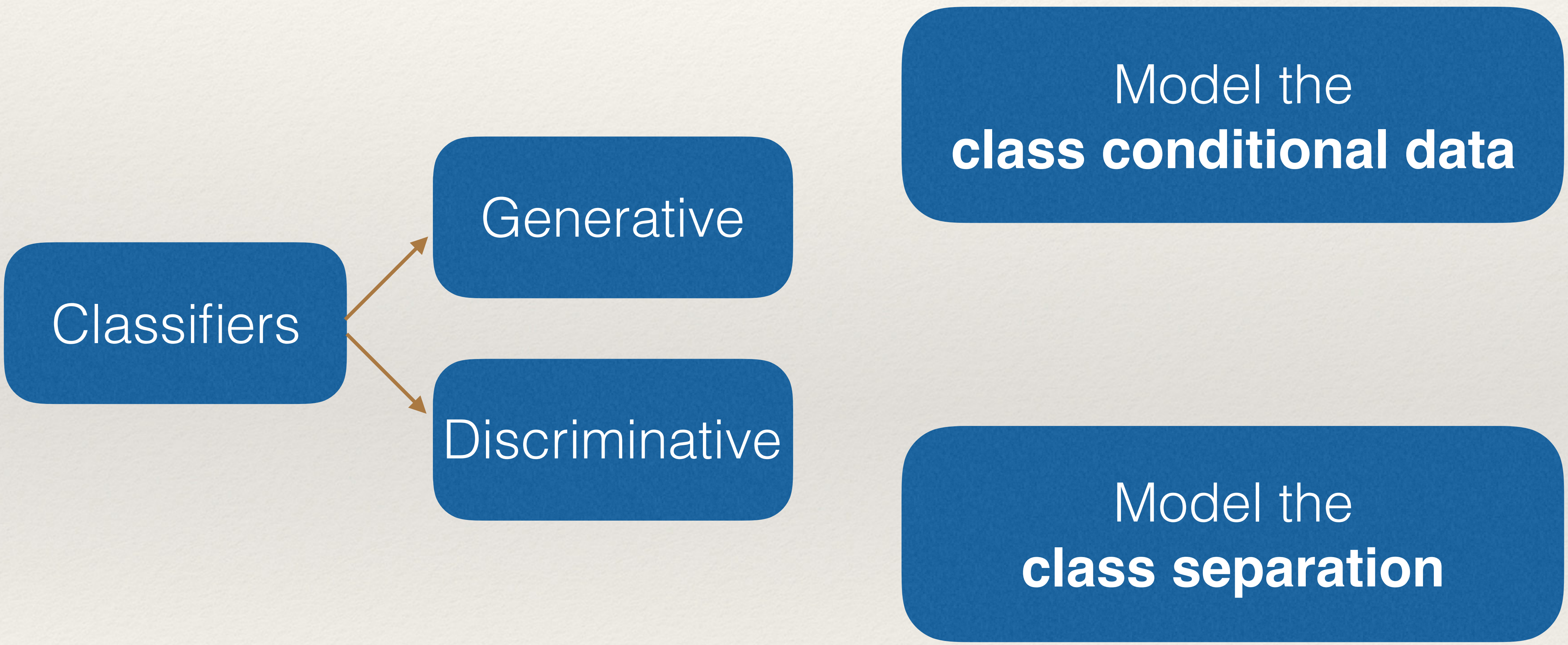
12-10-2016

Established
1911

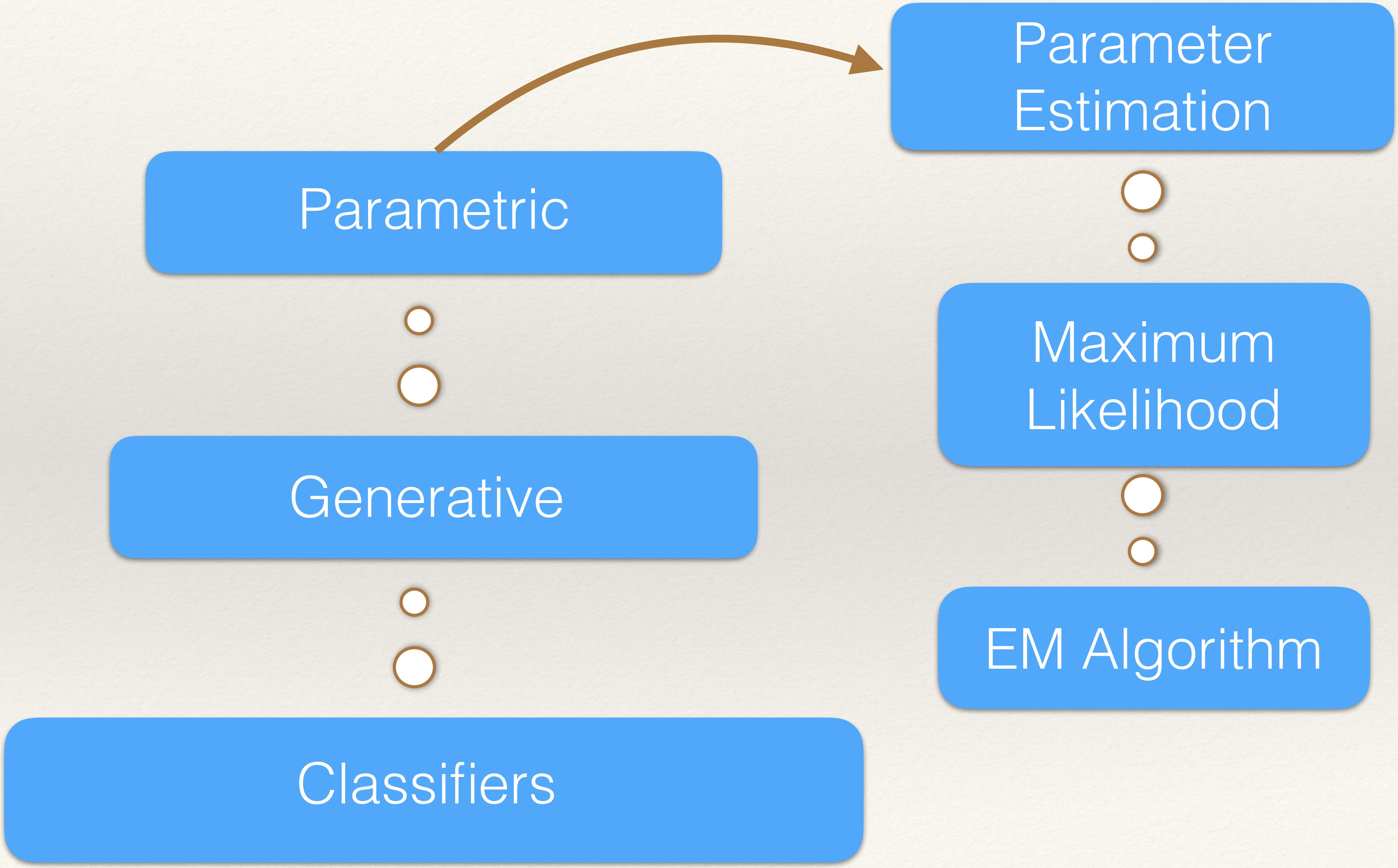
Classification Problems

- Inference problem - finding $p(C|\mathbf{x})$
- Decision problem - use the inference to classify
  - Maximum posterior rules.
- Approaches for inference estimation
  - Generative modeling - Determine likelihood $p(\mathbf{x}|C)$ and prior distribution $p(C)$
  - Discriminative modeling - directly model posterior

Classification Problems



Framework of Generative Models



Types of Generative Models

- Parametric Models
  - Collection of probability distributions which are described by a finite dimensional parameter set
- Gaussian model

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} exp\Big\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^* \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big\}$$

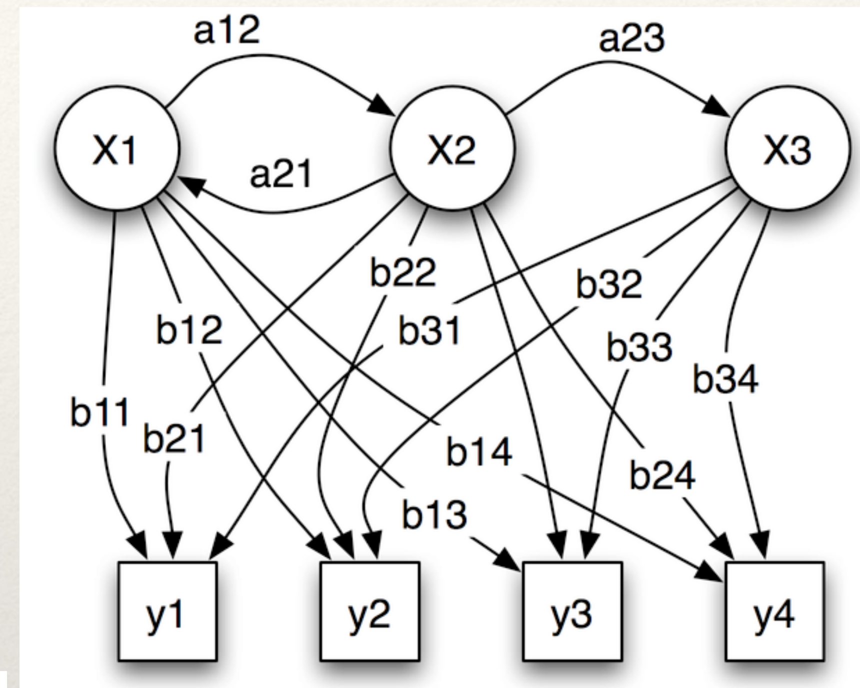
- Gaussian Mixture Model

$$p(\mathbf{x}|\boldsymbol{\Theta}) = \sum_{k=1}^{K} \alpha_k p(\mathbf{x}|\boldsymbol{\theta}_k)$$

$$p(\mathbf{x}|\boldsymbol{\theta}_k) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} exp\Big\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^* \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\Big\}$$

Types of Generative Models

- Hidden Markov Models



- An observation is a probabilistic function of a state, i.e., HMM is a doubly embedded stochastic process
- A DHMM is characterized by
  - N states S_j and M distinct observations v_k (alphabet size)
  - State transition probability distribution A
  - Observation symbol probability distribution B
  - Initial state distribution π

Parameter Estimation

- Given N data points $\{\mathbf{x}_i\}_{i=1}^{N}$ how do we estimate the parameters of model.
  - Several criteria can be used
  - The most popular method is the maximum likelihood estimation (MLE).

EM Algorithm

- Iterative procedure.
- Assume the existence of hidden variable associated with each data sample
- Let the current estimate (at iteration n) be Define the Q function as

$$Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^n) = E_{\mathbf{z}|\mathbf{X},\boldsymbol{\Theta}^n}\left[\log(P(\mathbf{X}, \mathbf{z}|\boldsymbol{\Theta})\right]$$

$$= \sum_{\mathbf{z}} \log(P(\mathbf{X}, \mathbf{z}|\boldsymbol{\Theta}))P(\mathbf{z}|\mathbf{X}, \boldsymbol{\Theta}^n)$$

EM Algorithm

- It can be proven that if we choose

$$\mathbf{\Theta}^{n+1} = arg \max_{\mathbf{\Theta}} Q(\mathbf{\Theta}, \mathbf{\Theta}^{n})$$

then $L(\mathbf{\Theta}^{n+1}) \geq L(\mathbf{\Theta}^{n})$

- In many cases, finding the maximum for the Q function may be easier than likelihood function w.r.t. the parameters.
- Solution is dependent on finding a good choice of the hidden variables which eases the computation
- Iteratively improve the log-likelihood function.

EM Summary

- Initialize with a set of model parameters (n=1)
- Compute the conditional expectation (E-step)

$$E_{\mathbf{z}|\mathbf{X},\mathbf{\Theta}^n}\left[\log(P(\mathbf{X},\mathbf{z}|\mathbf{\Theta}))\right]$$

- Maximize the conditional expectation w.r.t. parameter. (M-step) (n = n+1)
- Check for convergence
- Go back to E-step if model has not converged.

Applications of EM Algorithm

- GMM parameter estimation
  - hidden variables are alignment variables
  - estimating the weights, means, covariances
- HMM parameter estimation
  - hidden variables are state sequence (and alignment for GMM-HMMs).
  - estimating the transition probabilities and emission probabilities