Data Analytics
EEE 4774 & 6777

Module 4 - Classification

Generative Models

Spring 2022

Classification Problem & Approaches

Assumption:
Forward Model
$C_n \rightarrow \boxed{} \rightarrow x_n$

Testing

input
$\boldsymbol{x}_n$
e.g. animal img.

$\rightarrow$ Trained Classification Algorithm $\rightarrow$

Backward Model

output
$\hat{C}_n = [0,1,0, \dots, 0]$
{cat, dog, ..., duck}

1. Generative Models: (Likelihood, Prior, Posterior) modeling to first obtain likelihood and prior, and then posterior

Training: Tr. data $\rightarrow$ Model Fit (Param. Est) $\rightarrow$ Generative Model

$\max_{C_n} p(C_n|x_n) = \max_{C_n} p(x_n|C_n)\,p(C_n)$

$$p(C_n|\boldsymbol{x}_n) = \frac{\overbrace{p(\boldsymbol{x}_n|C_n)}^{\text{Likelihood}}\overbrace{p(C_n)}^{\text{prior}}}{p(\boldsymbol{x}_n)}$$

2. Discriminative Models: (Posterior) modeling to directly discriminate classes

posterior distribution $\rightarrow$

$$p(C_n|\boldsymbol{x}_n) = y(\boldsymbol{x}_n) = f(\boldsymbol{w}^T\boldsymbol{x}_n + w_0)$$

param.

Generative Models

$$p(C_n|\boldsymbol{x}_n) = \frac{p(\boldsymbol{x}_n|C_n)p(C_n)}{p(\boldsymbol{x}_n)}$$

$$\boxed{\max p(C_n|\boldsymbol{x}_n) = \max p(\boldsymbol{x}_n|C_n)\, p(C_n)}$$

Classific. decision using Generative models

e.g.,
Gaussian mixture model,
Naïve Bayes,

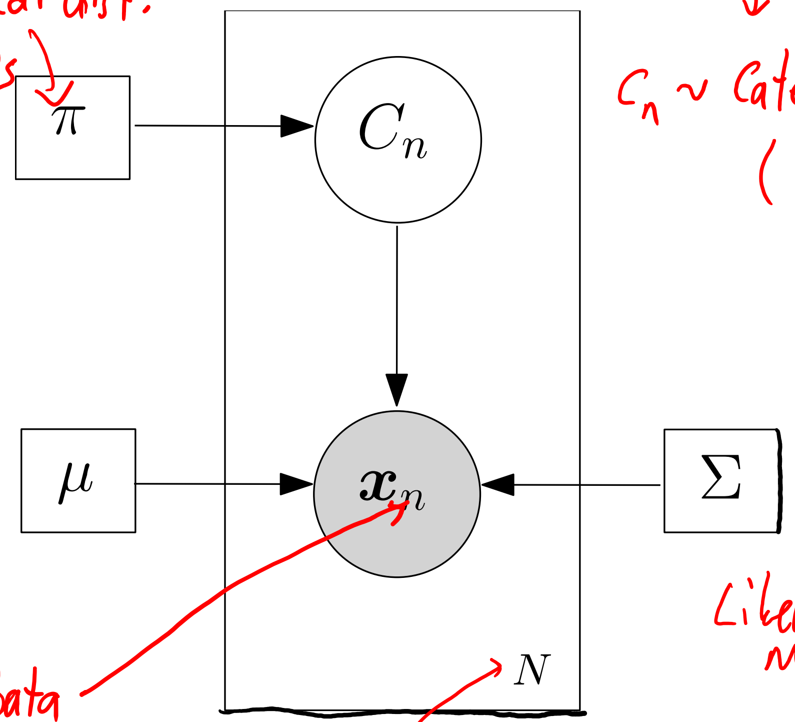
<u>Deep Generative Models:</u>
Variational autoencoder (VAE),
Generative adversarial network (GAN)

Graphical Model

Prior prob. of categorical dist. for class labels

Prior dist.

$C_n \sim \text{Categorical}(\pi)$



data index

plate notation shows repetition

Likelihood model

$x_n | C_n \sim \mathcal{N}(\mu, \Sigma)$

$x_n \in \mathbb{R}^d$

$\mu \in \mathbb{R}^d$

$\Sigma \in \mathbb{R}^{d \times d}$

Naïve Bayes

- Assumption: Conditioned on the class C_n input variables $x_{n1}, \ldots, x_{nd}$ are independent

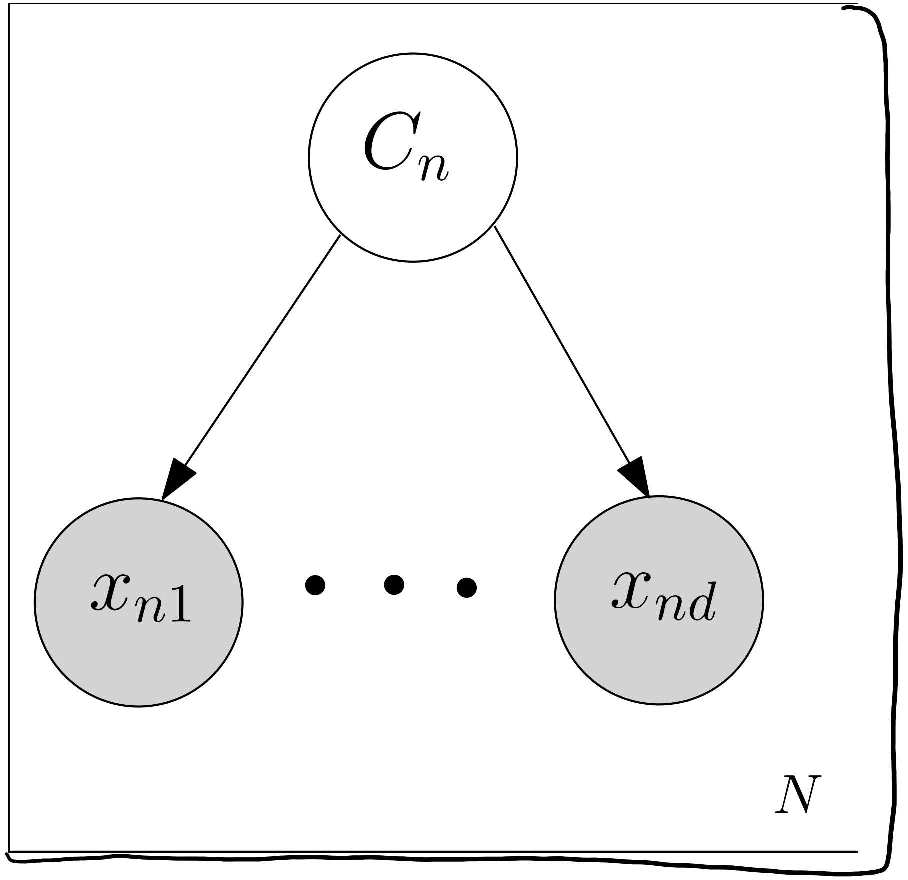
features

$$p(\boldsymbol{x}_n | C_n) = \prod_{i=1}^{d} p(x_{ni} | C_n)$$

$$p(\boldsymbol{x}_n) \neq \prod_{i=1}^{d} p(x_{ni})$$

- Simplified analysis: Univariate models instead of multivariate (esp. for large d)

- Limitation: Assumption does not hold in general



e.g. $x_{nd} \sim \mathcal{N}(\mu_d, \sigma_d^2)$

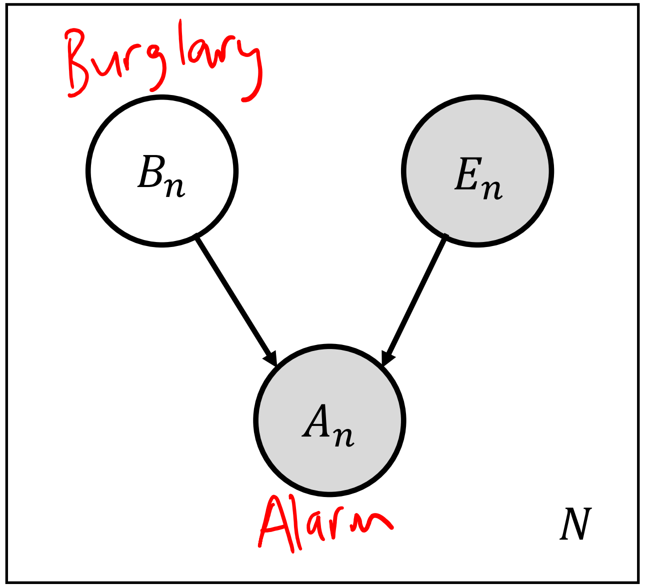
Example: Burglary-Alarm-Earthquake

Sally's burglar Alarm is sounding.
Has she been Burgled, or was the alarm triggered by an Earthquake?

Bayesian Model

↓ for

Causal relationship between observed and latent variables



Binary r.v.'s

↓

Bernoulli distrib.

Alarm, $A_n \in \{0,1\}$

$B_n \in \{0,1\}$

$E_n \in \{0,1\}$

Goal: Compute posterior $p(B \mid A)$ to decide whether $B=0$ or $B=1$

$p(B \mid A, E)$

Training & Model Building

Fit a Bernoulli distribution for each alarm case (4 cases) using training data

$$p(A|\underbrace{B,E}_{\text{given}})$$: likelihood model

from training data

Alarm = 1	Burglar	Earthquake
0.9999	1	1
0.99	1	0
0.99	0	1
0.0001	0	0

Prior probabilities for Burglary and Earthquake estimated from training data

$p(B = 1) = 0.01$ and $p(E = 1) = 0.000001$

Testing

$P(A=1|B=1,E=0)\ P(B=1,E=0)$

0.99 $\quad P(B=1)\ P(E=0)$

$0.01 \quad 1-0.000001$

observed data

$$p(B=1|\overbrace{A=1,E=0}^{\text{observed data}}) = \frac{p(B=1,A=1,E=0)}{\sum_B p(B,A=1,E=0)} \approx \underline{0.99}$$

$B=0$
$B=1$

$$p(B=1|A=1,E=1) = \frac{p(B=1,A=1,E=1)}{\sum_B p(B,A=1,E=1)} \approx \underline{0.01}$$

$$p(B=1|A=1) = p(B=1|A=1,E=1)p(E=1)$$
$$+p(B=1|A=1,E=0)p(E=0)$$
$$\approx \underline{0.99}$$

Earthquake explains away the fact that alarm is ringing.