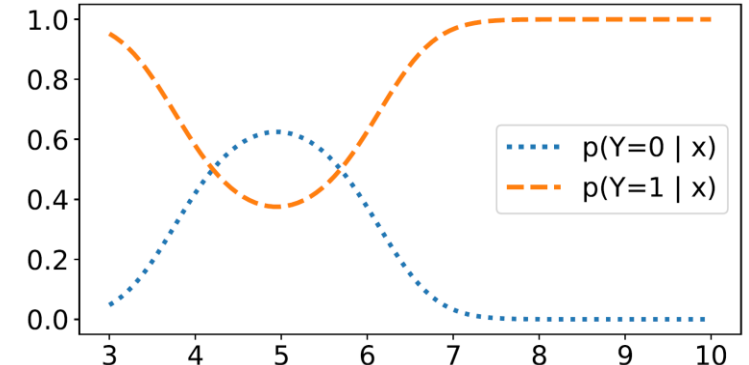
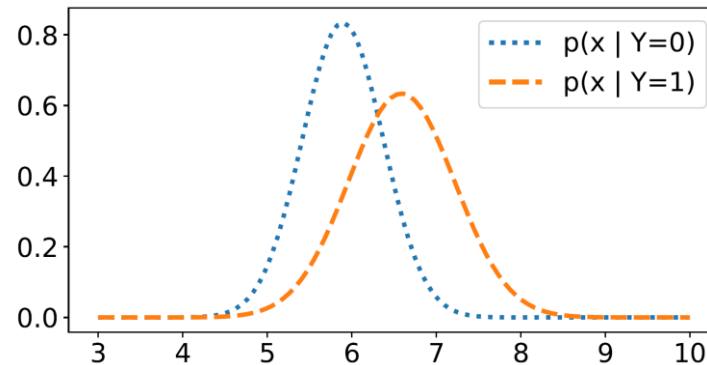
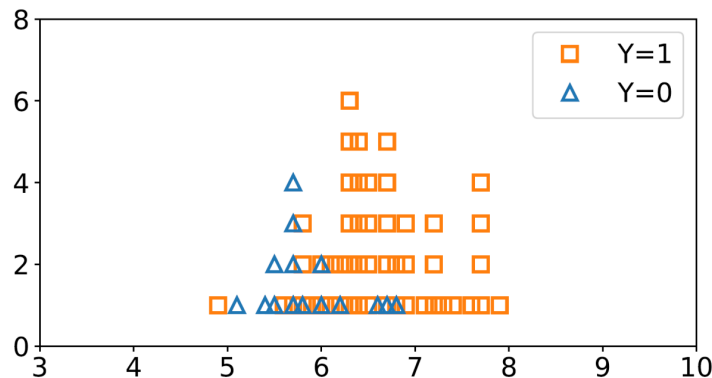


07-380 AI & ML II

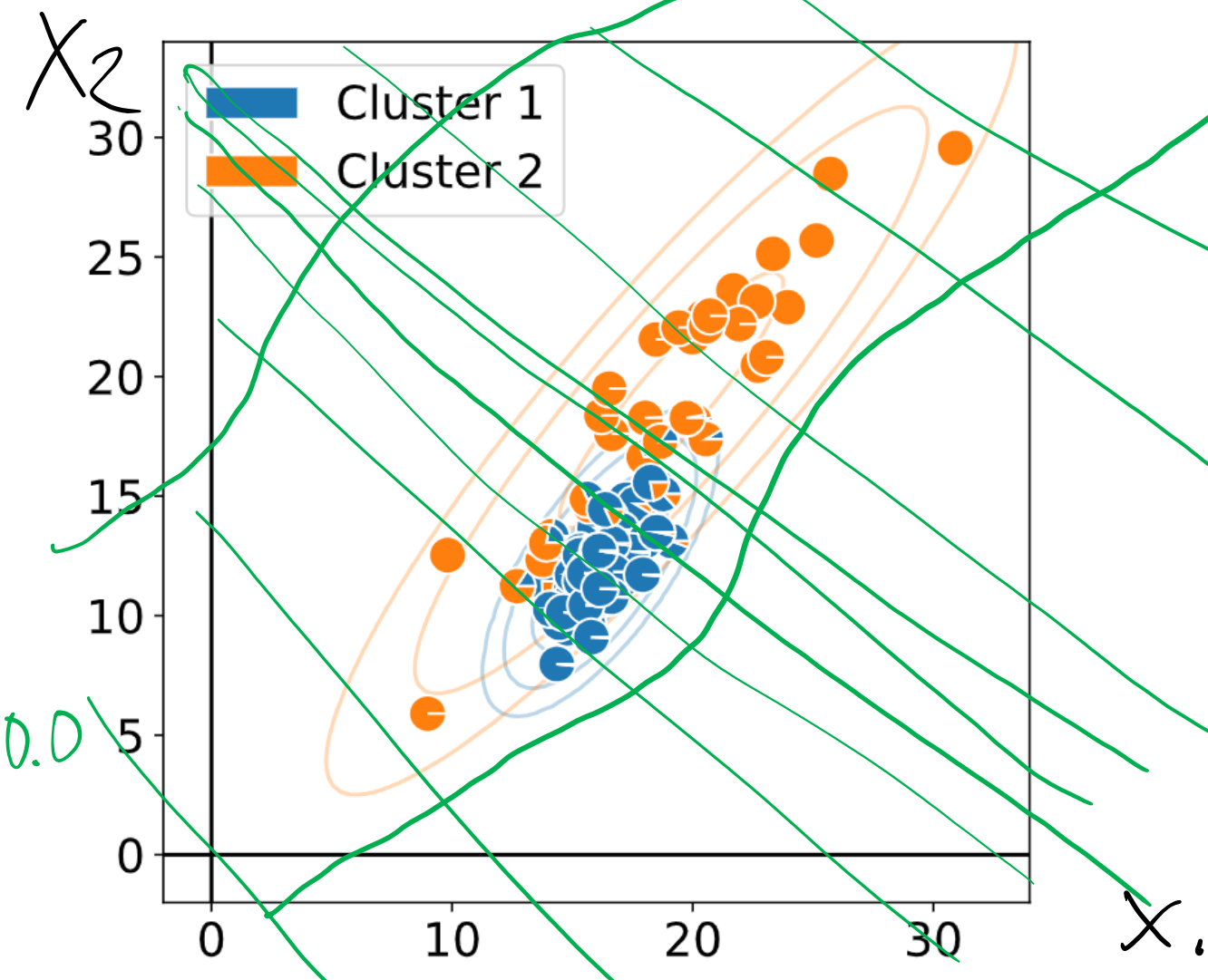
Probabilistic Generative Models: Discriminant Analysis & Naive Bayes



Instructors: Pat Virtue & Mohammad Salameh

Classifying Cats and Dogs

Logistic Regression??



$$g(w_1 x_1 + w_2 x_2 + b)$$

Bayes Rule (again)

Two Applications of Bayes Rule

$$p(a | b) = \frac{p(b | a) p(a)}{p(b)}$$

$$p(\theta | \mathcal{D}) = \frac{p(\mathcal{D} | \theta) p(\theta)}{p(\mathcal{D})}$$

Data and Parameters

M(LE)

$$p(y | x) = \frac{p(x | y) p(y)}{p(x)}$$

ML Input and Output Data

New today

Bayes Rule: Data and Parameters

Likelihood and Prior

$$p(\theta \mid \mathcal{D}) = \frac{p(\theta) p(\mathcal{D} \mid \theta)}{p(\mathcal{D})}$$

$$p(\theta \mid \mathcal{D}) \propto p(\theta) p(\mathcal{D} \mid \theta)$$

$$= p(\theta) \prod_{i=1}^N p(y^{(i)} \mid \theta)$$

$p(\mathcal{D} \mid \theta)$ is the **likelihood**,

$p(\theta)$ is the **prior** on the parameters

$p(\theta \mid \mathcal{D})$ is the **posterior**

(Conditional) Likelihood and Prior

$$p(\theta) \prod p(y^{(i)} \mid x^{(i)} \mid \theta)$$

Bayes Rule: ML Input and Output Data

Class Conditional and Class Prior

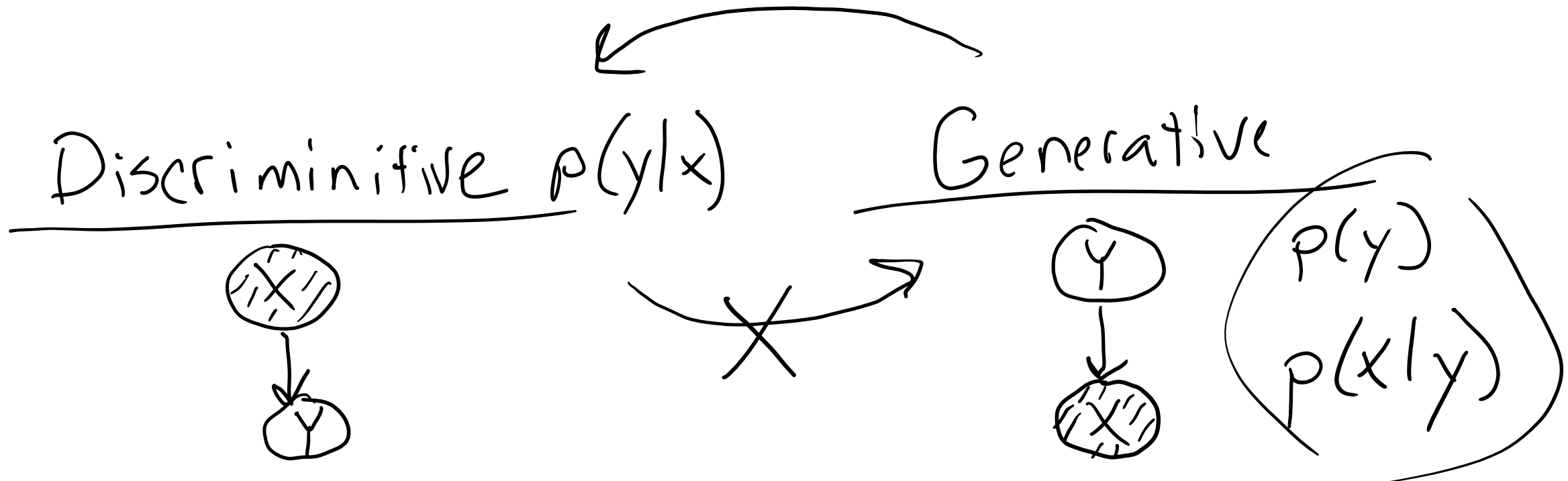
$$p(y | x) = \frac{p(x | y) p(y)}{p(x)}$$

$$p(y | x) \propto p(x | y) p(y)$$

$p(x | y)$ is defined as the class conditional distribution

$p(y)$ is the label class prior distribution

$p(y | x)$ is the conditional likelihood



Bayes Rule

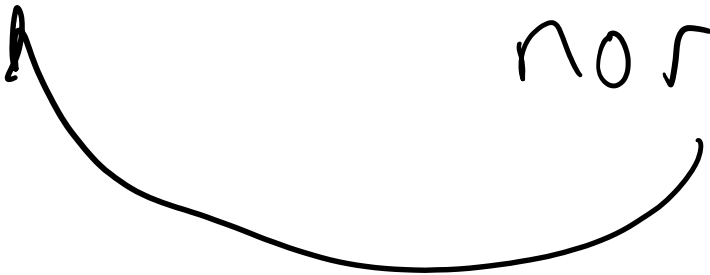
Where did the parameters go?!?

Gaussian

Bernoulli

Hiding

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

$$p(y|x) = \frac{p(x|y, \underline{\mu}, \underline{\sigma^2}) p(y|\underline{\phi})}{\text{normalizing}}$$


Discriminative vs Generative Models

Iris classification

Optimization: Generative vs Discriminative

Discriminative: model $p(y | x, \theta)$ directly

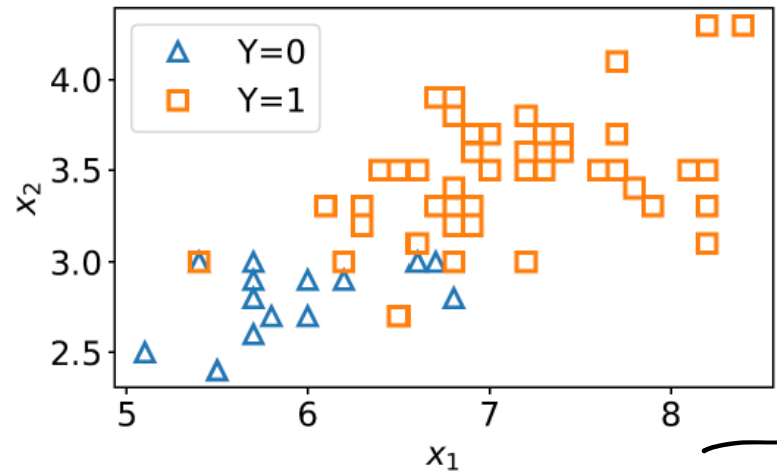
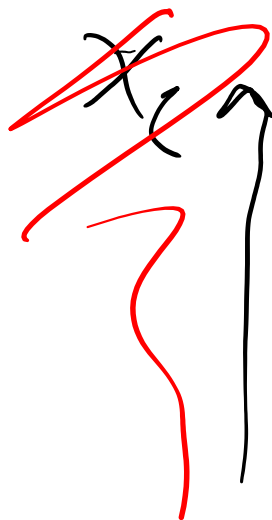
- Learn parameters θ from data

Generative: model $p(y | \theta_{class})$ and $p(x | y, \theta_{class\ conditional})$

- Learn parameters θ_{class} and $\theta_{class\ conditional}$ from data
- Use Bayes rule to compute $p(y | x, \theta_{class}, \theta_{class\ conditional})$

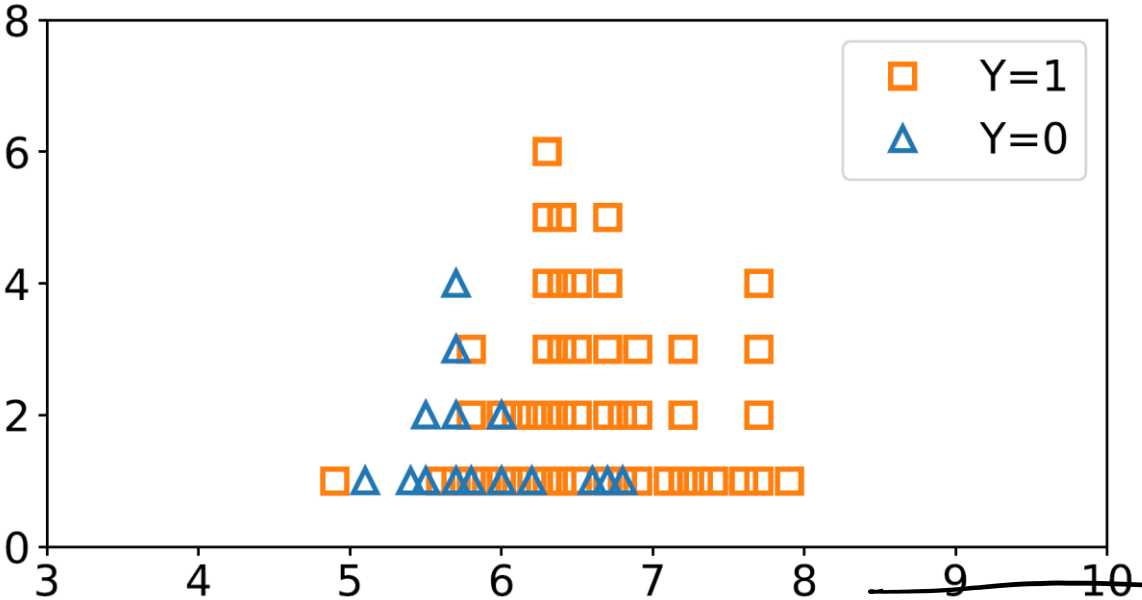
$$p(y | x, \theta_{class}, \theta_{class\ conditional}) \propto p(x | y, \theta_{class\ conditional}) p(y | \theta_{class})$$

Iris Classification



x_1 $\rightarrow$ x_1

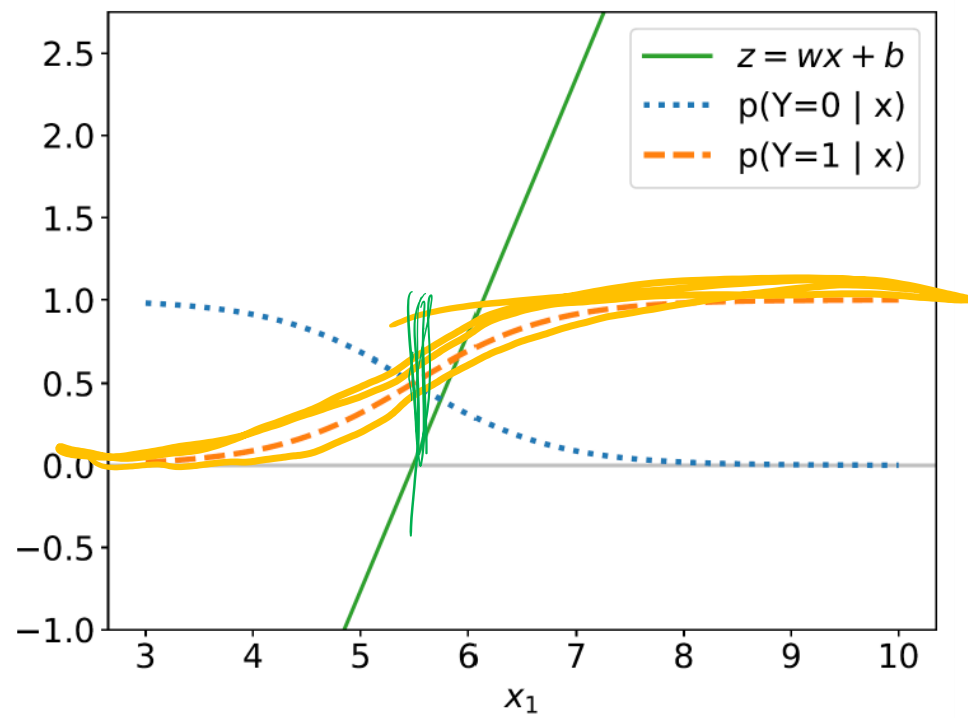
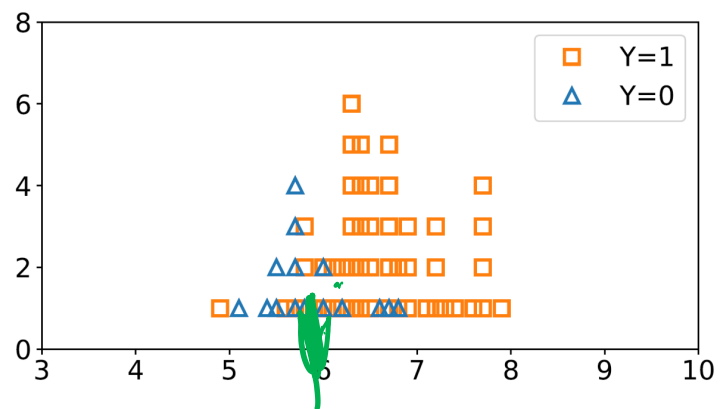
Histogram



x_1 $\rightarrow$

Iris Classification

Logistic Regression??



Iris Classification

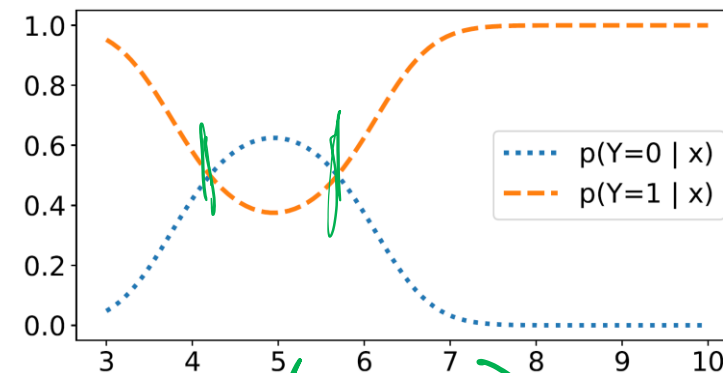
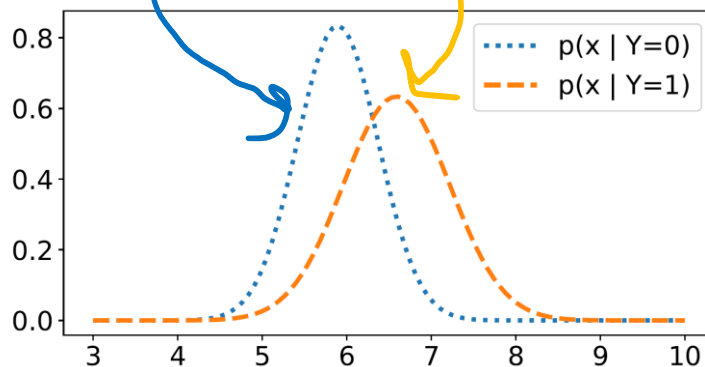
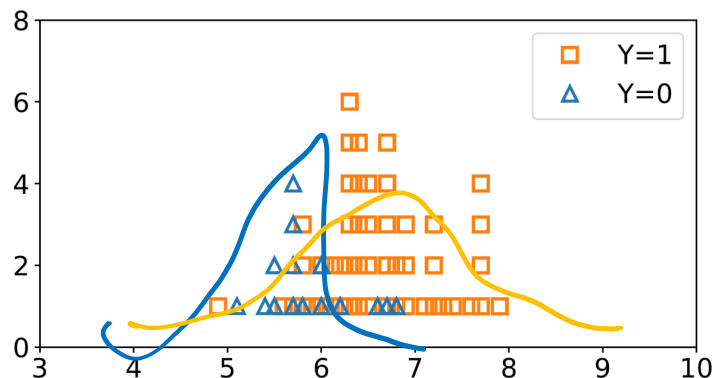
$$P(Y=0)$$

$$P(Y=1)$$

Generative

$$P(x | Y=0, \mu_{Y=0}, \sigma_{Y=0}^2)$$

$$P(x | Y=1, \mu_{Y=1}, \sigma_{Y=1}^2)$$

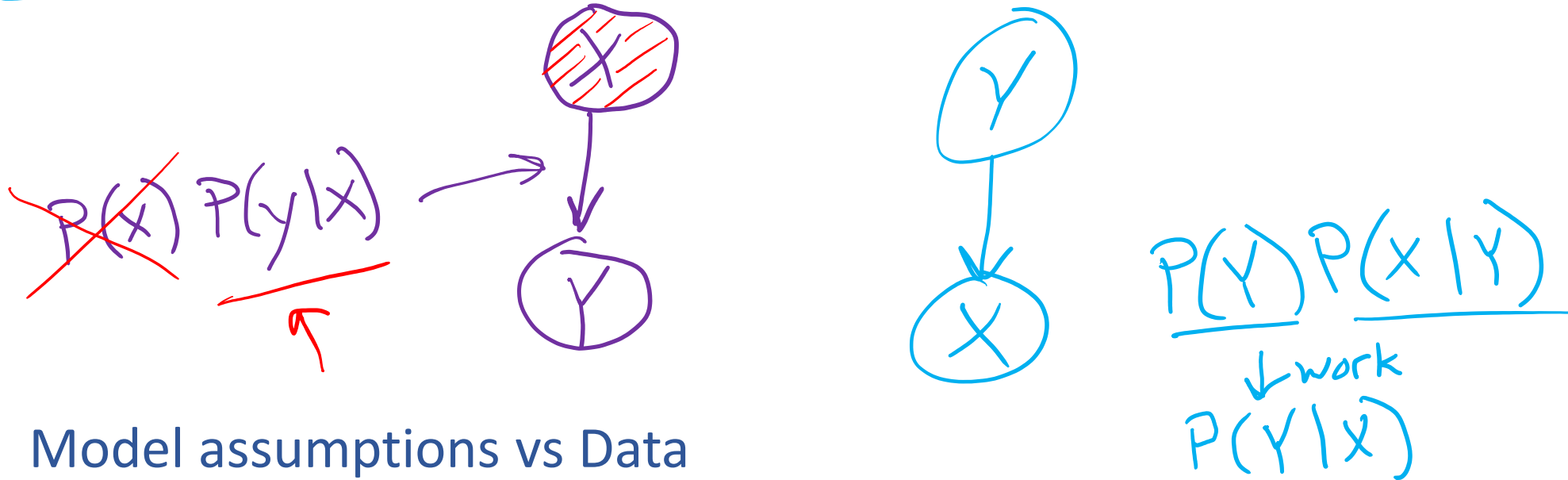


$p(y | x)$

Generative vs Discriminative Modeling

Discriminative: $p(y | x)$

Generative: $p(y | x) = \alpha p(x, y) = \alpha p(x | y) p(y)$



Model assumptions vs Data

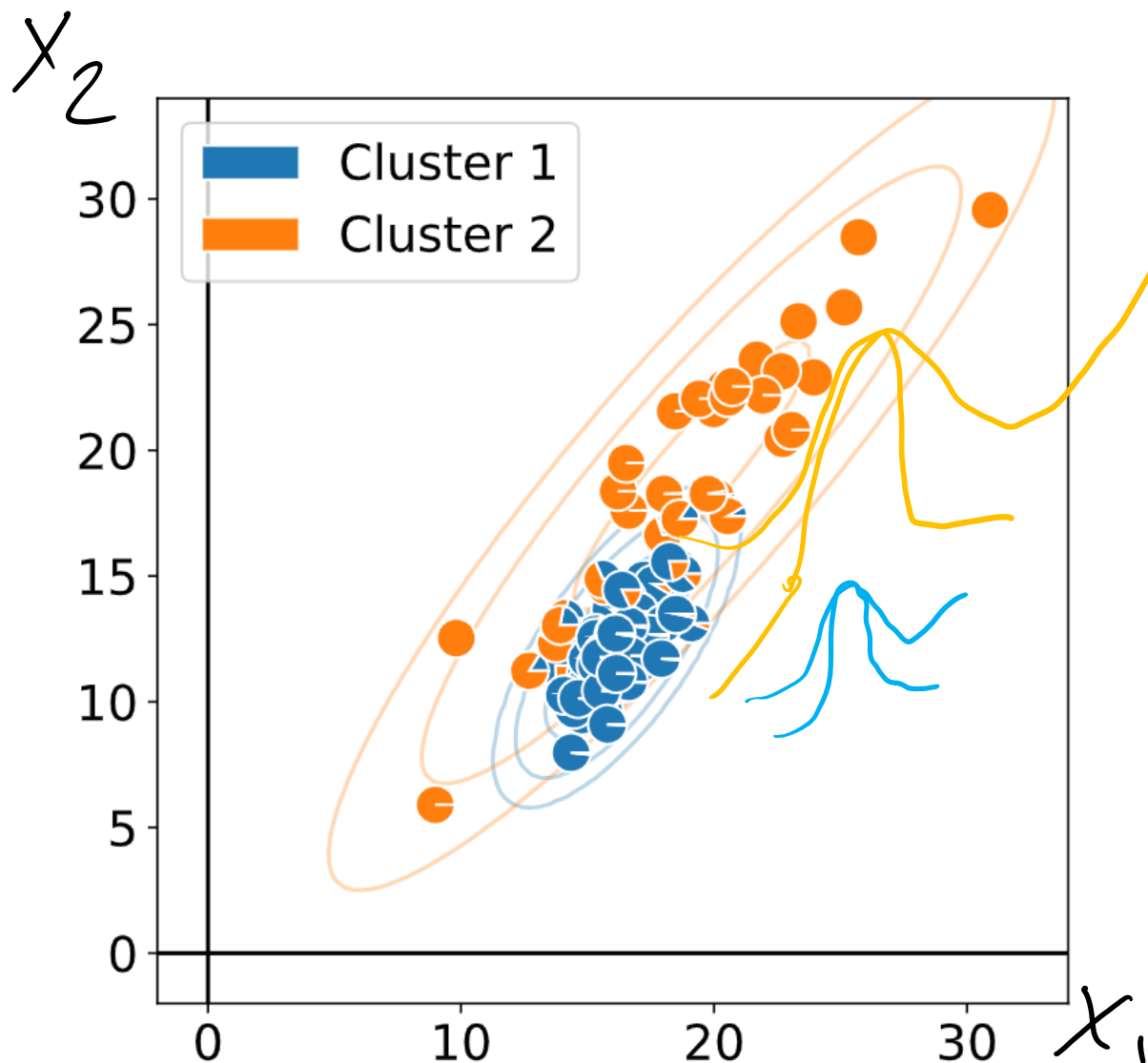
- Discriminative: **Weak** model assumption, relies on **lots of data**
- Generative: **Stronger** model assumption, requires **less data**

Discriminative and Generative

Formulation for optimization

Estimating Cats and Dogs

params
11



Bernoulli + Gaussian

$$p(Y = \text{cat} | \phi) = (1 - \phi)$$

$$p(Y = \text{dog} | \phi) = \phi$$

$$P(x_1, x_2 | Y = \text{cat}) \sim \mathcal{N}(\vec{\mu}_{\text{cat}}, \Sigma_{\text{cat}})$$

$$P(x_1, x_2 | Y = \text{dog}) \sim \mathcal{N}(\vec{\mu}_{\text{dog}}, \Sigma_{\text{dog}})$$

Handwritten notes and arrows indicate the parameters of the Gaussian distributions:

- Red arrow 1 points to ϕ in the Bernoulli equation.
- Red arrow 2 points to $\vec{\mu}_{\text{cat}}$ in the Gaussian equation for cats.
- Red arrow 3 points to Σ_{cat} in the Gaussian equation for cats.
- Red arrow 2 points to $\vec{\mu}_{\text{dog}}$ in the Gaussian equation for dogs.
- Red arrow 3 points to Σ_{dog} in the Gaussian equation for dogs.

Reminder: Likelihood

Simple example with three i.i.d. samples of $Y \sim \text{Bernoulli}(\phi)$

$$\mathcal{D} = \{y^{(i)}\}_{i=1}^3$$

Likelihood

$$p(\mathcal{D}; \theta)$$

$$= p(y^{(1)}, y^{(2)}, y^{(3)}; \phi^{(1)}, \phi^{(2)}, \phi^{(3)})$$
 Expand notation

$$= p(y^{(1)}, y^{(2)}, y^{(3)}; \phi)$$
 Identically distributed

$$= p(y^{(1)}; \phi) p(y^{(2)}; \phi) p(y^{(3)}; \phi)$$
 Independent

$$= \prod_{i=1}^3 p(y^{(i)}; \phi) \sim \text{Bern}(\phi)$$

Generative MLE

$$L(\phi, \Theta) = p(\mathcal{D} \mid \phi, \Theta)$$

$$= \prod_{n=1}^N p(\mathcal{D}^{(n)} \mid \phi, \Theta) \quad \text{i.i.d assumption}$$

$$= \prod_{n=1}^N p(y^{(n)}, \mathbf{x}^{(n)} \mid \phi, \Theta)$$

Generative model

$$= \prod_{n=1}^N p(y^{(n)} \mid \phi) p(\mathbf{x}^{(n)} \mid y^{(n)}, \Theta)$$

Generative model story

$$= \prod_{n=1}^N p(y^{(n)} \mid \phi) p(x_1^{(n)}, x_2^{(n)}, \dots, x_M^{(n)} \mid y^{(n)}, \Theta)$$

Gaussian

Gaussian

$$p(x_1 \mid y) p(x_2 \mid y) p(x_3 \mid y)$$

spam

Not spam

Discriminative
MLE
 $\prod_{n=1}^N P(y^{(n)} \mid \tilde{\mathbf{x}}^{(n)}, \tilde{\Theta})$

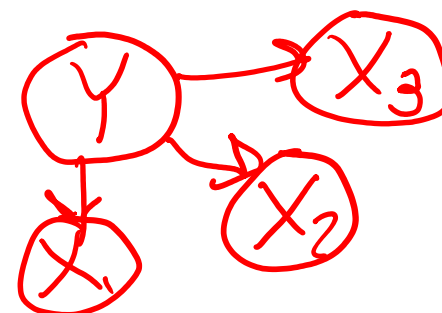
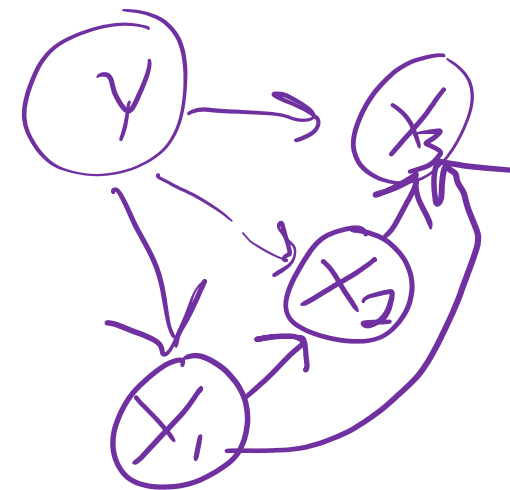
$$\mathcal{D} = \{y^{(i)}, x^{(i)}\}_{i=1}^N$$

$$y^{(i)} \in \{0, 1\}$$

$$\mathbf{x}^{(i)} \in \{0, 1\}^M$$

$$\phi \in [0, 1]$$

$$\Theta \in [0, 1]^{M \times 2}$$



Generative + MAP

MLE vs MAP vs Generative vs Discriminative

Maximum likelihood estimation

Maximum *a posteriori* estimation

Discriminative

Models only the
conditional
likelihood,
 $\prod p(y | x, \theta)$

$$p(\mathcal{D}^* | \theta) \quad \text{* conditional likelihood}$$

$$= \prod_{i=1}^N p(y^{(i)} | x^{(i)}, \theta)$$

$$p(\theta) p(\mathcal{D}^* | \theta) \quad \text{* conditional likelihood}$$

$$= p(\theta) \prod_{i=1}^N p(y^{(i)} | x^{(i)}, \theta)$$

Generative

Models the full
joint likelihood,
 $\prod p(x, y | \theta)$

$$p(\mathcal{D} | \theta) \quad \text{* actual likelihood}$$

$$= \prod_{i=1}^N p(x^{(i)}, y^{(i)} | \theta)$$

$$= \prod_{i=1}^N \underbrace{p(x^{(i)} | y^{(i)}, \theta)}_{\text{class-conditional}} \underbrace{p(y^{(i)} | \theta)}_{\text{class prior}}$$

$$p(\theta) p(\mathcal{D} | \theta) \quad \text{* actual likelihood}$$

$$= p(\theta) \prod_{i=1}^N p(x^{(i)}, y^{(i)} | \theta)$$

$$= p(\theta) \prod_{i=1}^N \underbrace{p(x^{(i)} | y^{(i)}, \theta)}_{\text{class-conditional}} \underbrace{p(y^{(i)} | \theta)}_{\text{class prior}}$$

Naïve Bayes

Exercise: SPAM Classification

Recitation Exercise

Training Data

Spam?

1

E-mail body

Money is free now

0

Pat teach 380

0

Pat free to teach

1

Sir money to teach

1

Pat free money now

0

Teach 380 now

0

Pat to teach 380

Vocabulary

380

free

is

money

now

Pat

Sir

teach

to

tomorrow

Worksheet ☺

Go ahead and start
- Help each other

Joint

$$P(Y = 0, X_1, \dots, X_M)$$

$$P(Y = 1, X_1, \dots, X_M)$$

$$0.1$$

$$0$$

$$P(Y = 0 | X_1, \dots, X_M)$$

$$P(Y = 1 | X_1, \dots, X_M)$$

$$\frac{0.1}{0.1+0} = 1$$

$$\frac{0}{0.1+0} = 0$$

$$P(Y=0 | \vec{x}) = \frac{P(Y=0, \vec{x})}{\sum_{y=0} P(Y=y, \vec{x})}$$

$$\sum_{y=0} P(Y=y, \vec{x})$$

normalization

Conditional Independence and Naïve Bayes

Independence

$$A \perp\!\!\!\perp B$$

$$P(A | B) = P(A)$$

$$P(B | A) = P(B)$$

$$P(A, B) = P(A)P(B)$$

Conditional independence

$$A \perp\!\!\!\perp B \mid C$$

$$P(A | B, \underline{C}) = P(A | \underline{C})$$

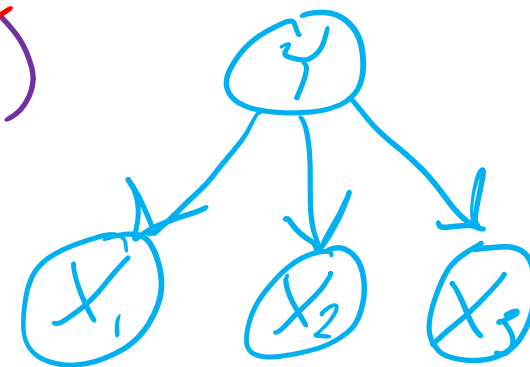
$$P(B | A, \underline{C}) = P(B | \underline{C})$$

$$P(A, B | \underline{C}) = P(A | \underline{C})P(B | \underline{C})$$

Naïve Bayes assumption

$$\forall j \neq k \quad X_j \perp\!\!\!\perp X_k \mid Y$$

$$P(X_2 | Y, \cancel{X_1}) \Rightarrow P(X_2 | Y)$$



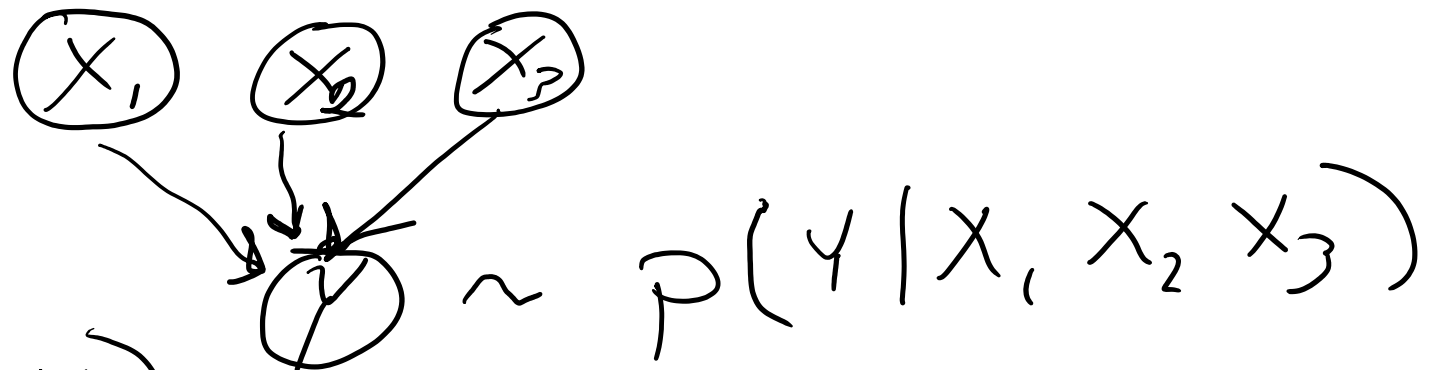
Generative Stories

sampling
or
creating

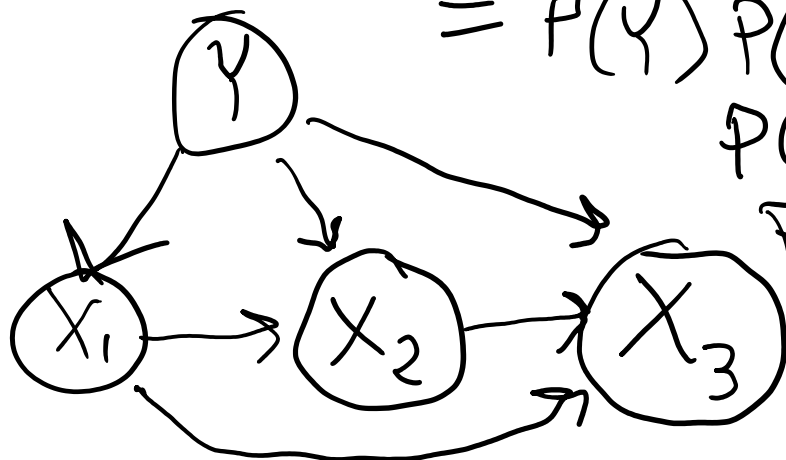
Generative Story Examples

Spam Generation

Discriminative

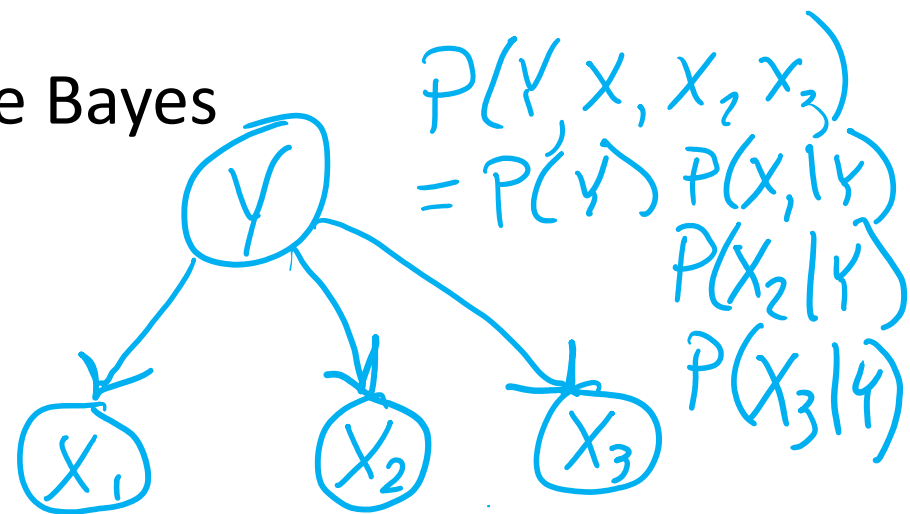


Generative



$$\begin{aligned} P(Y, X_1, X_2, X_3) \\ = P(Y) P(X_1 | Y) \\ P(X_2 | Y, X_1) \\ P(X_3 | Y, X_1, X_2) \\ \text{Chain rule} \end{aligned}$$

Generative + Naive Bayes



$$\begin{aligned} P(Y, X_1, X_2, X_3) \\ = P(Y) P(X_1 | Y) \\ P(X_2 | Y) \\ P(X_3 | Y) \end{aligned}$$

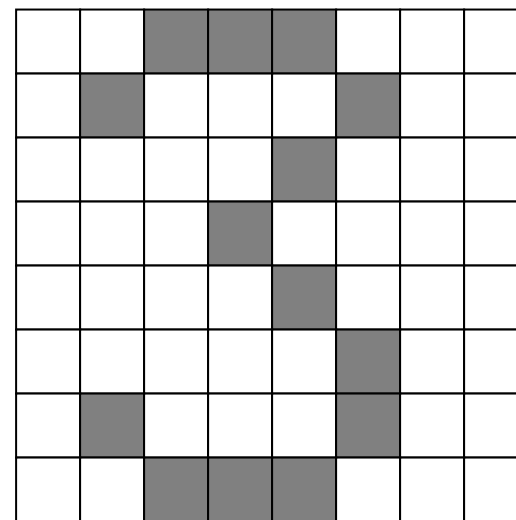
Generative Story Examples

Hand-written digits

y: Digit class: 0-9

x: Pixels in images

Discriminative



$x_1, \dots, x_{64}$

$P(Y, x_1, \dots, x_{64})$

Generative

← Size of joint?

Generative + Naive Bayes

10×2^{64}

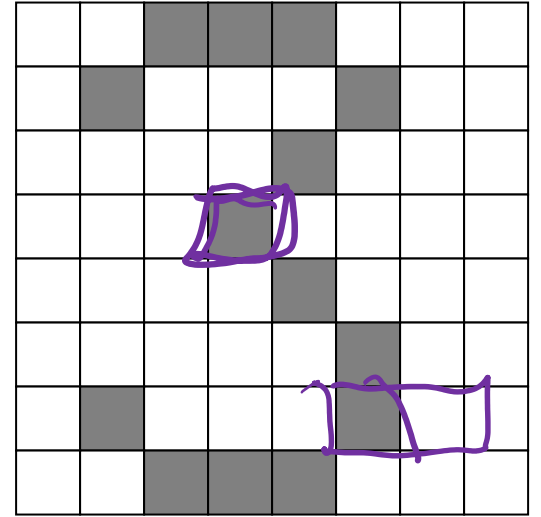
tiny
 2^{64}

$P(\vec{x}) \leftarrow 2^{64}$

Multivariate Generative Models

Hand-written digits: How many parameters?

- $P(Y)$
- $P(\mathbf{X} \mid Y = 3)$
 $= p(X_1, X_2, \dots, X_{64} \mid Y = 3)$

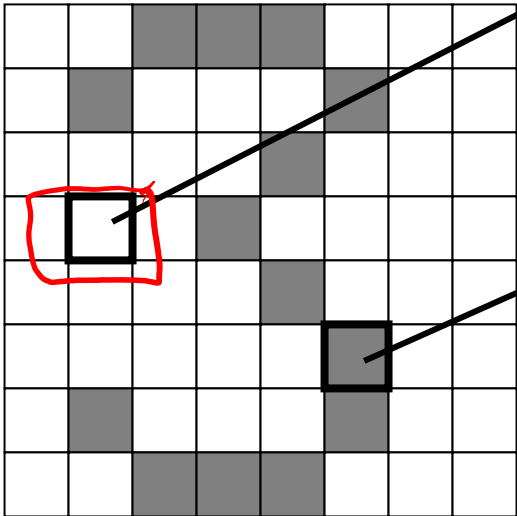


Naïve Bayes assumption (bag of pixels)

- $P(Y)$
- $P(\mathbf{X} \mid Y = 3)$
 $= p(X_1 \mid Y = 3)p(X_2 \mid Y = 3) \dots p(X_{64} \mid Y = 3)$

Naïve Bayes for Digits

y	$P(Y)$
1	0.1
2	0.1
3	0.1
4	0.1
5	0.1
6	0.1
7	0.1
8	0.1
9	0.1
0	0.1



y	$P(X_{3,1} = 1 \mid y)$
1	0.01
2	0.05
3	0.05
4	0.30
5	0.80
6	0.90
7	0.05
8	0.60
9	0.50
0	0.80

y	$P(X_{5,5} = 1 \mid y)$
1	0.05
2	0.01
3	0.90
4	0.80
5	0.90
6	0.90
7	0.25
8	0.85
9	0.60
0	0.80

$$\phi_{y=3, x_{3,1}}$$

Naïve Bayes MLE

$$L(\phi, \Theta) = p(\mathcal{D} \mid \phi, \Theta)$$

$$= \prod_{n=1}^N p(\mathcal{D}^{(n)} \mid \phi, \Theta) \quad \text{i.i.d assumption}$$

$$= \prod_{n=1}^N p(y^{(n)}, \mathbf{x}^{(n)} \mid \phi, \Theta)$$

$$= \prod_{n=1}^N p(y^{(n)} \mid \phi) p(\mathbf{x}^{(n)} \mid y^{(n)}, \Theta) \quad \text{Generative model}$$

$$= \prod_{n=1}^N p(y^{(n)} \mid \phi) p(x_1^{(n)}, x_2^{(n)}, \dots, x_M^{(n)} \mid y^{(n)}, \Theta)$$

$$= \prod_{n=1}^N p(y^{(n)} \mid \phi) \prod_{m=1}^M p(\underline{x_m^{(n)}} \mid y^{(n)}, \theta_{m,y}) \quad \text{Naïve Bayes}$$

$$\begin{aligned} \mathcal{D} &= \{y^{(n)}, \mathbf{x}^{(n)}\}_{n=1}^N \\ y^{(n)} &\in \{0,1\} \\ \mathbf{x}^{(n)} &\in \{0,1\}^M \\ \phi &\in [0,1] \\ \Theta &\in [0,1]^{M \times 2} \end{aligned}$$



Generative Models

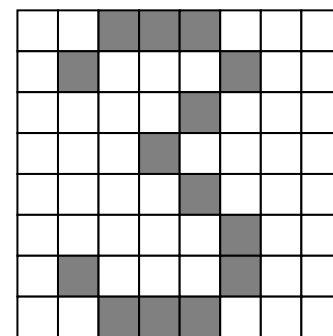
SPAM:

- Class distribution: $Y \sim \text{Bern}(\phi)$
- Class conditional distribution: $X_m \sim \text{Bern}(\theta_{m,y})$
- Naïve Bayes X_i conditionally independent X_j given Y for all $i \neq j$
$$p(X_i, X_j | Y) = p(X_i | Y) | p(X_j | Y)$$

Digits:

- Class distribution: $Y \sim \text{Categorical}(\boldsymbol{\phi})$
- Class conditional distribution: $X_m \sim \text{Bern}(\theta_{m,y})$
- Naïve Bayes X_i conditionally independent X_j given Y for all $i \neq j$
$$p(X_i, X_j | Y) = p(X_i | Y) | p(X_j | Y)$$

✓ $[\phi_0 \phi_1 \phi_2 \phi_3 \dots \phi_9]$



Generative Models with Continuous Features

Iris dataset:

- Class distribution: $Y \sim \text{Bern}(\phi)$
- Class conditional distribution: Multivariate Gaussian $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$
- Naïve Bayes assumption?

Poll 1

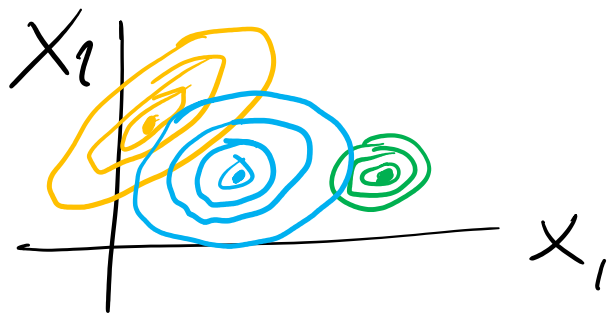
Iris dataset:

- Class distribution: $Y \sim \text{Categorical}(\boldsymbol{\phi})$
- Class conditional distribution: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$
- Three classes
- Using two input features $\mathbf{x} \in \mathbb{R}^2$

How many parameters are in our model?

- | | |
|------|--------|
| A. 3 | E. 15 |
| B. 6 | F. 18 |
| C. 7 | G. 920 |
| D. 9 | |

Poll 1



Iris dataset:

- Class distribution: $Y \sim \text{Categorical}(\phi)$
- Class conditional distribution: $\mathbf{X} \sim \mathcal{N}(\mu_y, \Sigma_y)$
- Three classes **A B C**
- Using two input features $\mathbf{x} \in \mathbb{R}^2$

3 categorical params
 $\phi_A \phi_B \phi_C$, s.t. $\sum \phi_k = 1$

$$P(\mathbf{x} | Y=A) \sim \mathcal{N}(\vec{\mu}_A, \Sigma_A)$$

$$P(\mathbf{x} | Y=B) \sim \mathcal{N}(\vec{\mu}_B, \Sigma_B)$$

$$p(\mathbf{x} | Y=C) \sim \mathcal{N}(\vec{\mu}_C, \Sigma_C)$$

How many parameters are in our model?

- A. 3
- B. 6
- C. 7
- D. 9

E. 15

F. 18

G. 920

$$3 + 15$$

$$3 \times \text{for } 3 \text{ classes} \\ = 3(2+3) = 15$$

2D Gaussian
→ 2 + 3 params

$$\begin{bmatrix} \mu_{c1} \\ \mu_{c2} \end{bmatrix}, \begin{bmatrix} \sigma_{c11} & \sigma_{c12} \\ \sigma_{c12} & \sigma_{c22} \end{bmatrix}$$

symmetric

Poll 2

Iris dataset:

- Class distribution: $Y \sim \text{Bern}(\phi)$
- Class conditional distribution: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$
- Naïve Bayes assumption?

Which of the following pairs of Gaussian class conditional distributions satisfy the Naïve Bayes assumptions? Select ALL that apply.

A. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

B. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

C. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$

D. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Poll 2

Iris dataset:

- Class distribution: $Y \sim \text{Bern}(\phi)$
- Class conditional distribution: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$
- Naïve Bayes assumption?

Which of the following pairs of Gaussian class conditional distributions satisfy the Naïve Bayes assumptions? Select ALL that apply.

A. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ✓ ✓

B. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ ✓ ✓

C. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$ ✗ ✗

D. $\boldsymbol{\mu}_{y=0} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=0} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \boldsymbol{\mu}_{y=1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{y=1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ✗ ✓

Class-conditional Gaussian Distributions

Iris dataset:

- Class distribution: $Y \sim \text{Bern}(\phi)$ (or Categorical)
- Class conditional distribution: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$

- Naïve Bayes assumption: diagonal $\boldsymbol{\Sigma}_y$'s (if $x_i \perp x_k | y$
then $\text{cov}(x_i, x_k) = 0$
→ off-diagonals of $\boldsymbol{\Sigma}$ are zero)
- Linear Decision Boundary: if covariance is the same for each class
 $\boldsymbol{\Sigma}_{Y=0} = \boldsymbol{\Sigma}_{Y=1}$
- Quadratic Decision Boundary: technically a cross-section of a quadratic function if $\boldsymbol{\Sigma}_{Y=0} \neq \boldsymbol{\Sigma}_{Y=1}$

Additional Slides

Poll

Let's say we gave an exam that had 3 different versions where the probability of different exams was modeled as a **categorical** random variable Y and the scores from each exam were modeled as $X_{Y=k} \sim \mathcal{N}(\mu_k, \sigma_k^2)$. $\mathcal{D} = \{x^{(i)}, y^{(i)}\}_{i=1}^N$

How many parameters are in this **generative** model?

- A. 2
- B. 3
- C. 6
- D. 9
- E. $3N$

Poll

Let's say we gave an exam that had 3 different versions where the probability of different exams Y given the scores from the exam x was modeled as a **multiclass logistic regression** $p(Y_k = 1 \mid x)$ trained on dataset $\mathcal{D} = \{x^{(i)}, y^{(i)}\}_{i=1}^N$

How many parameters are in this **discriminative** model?

- A. 2
- B. 3
- C. 6
- D. 9
- E. $3N$

Poll

Which of the following probability models will allow us to sample new data, $\mathcal{D} = \{x_1^{(i)}, x_2^{(i)}, y^{(i)}\}_{i=1}^N$ where x_1, x_2, y are length, weight, and species (cat or dog), respectively? Assume that we have learned all of the corresponding parameters

A) Logistic regression

$$p(Y = cat \mid x_1, x_2, w_1, w_2, b)$$

B) Bernoulli + Gaussian

$$p(Y = cat \mid \phi)$$

$$p(x_1 \mid cat, \mu_{cat,1}, \sigma_{cat,1})$$

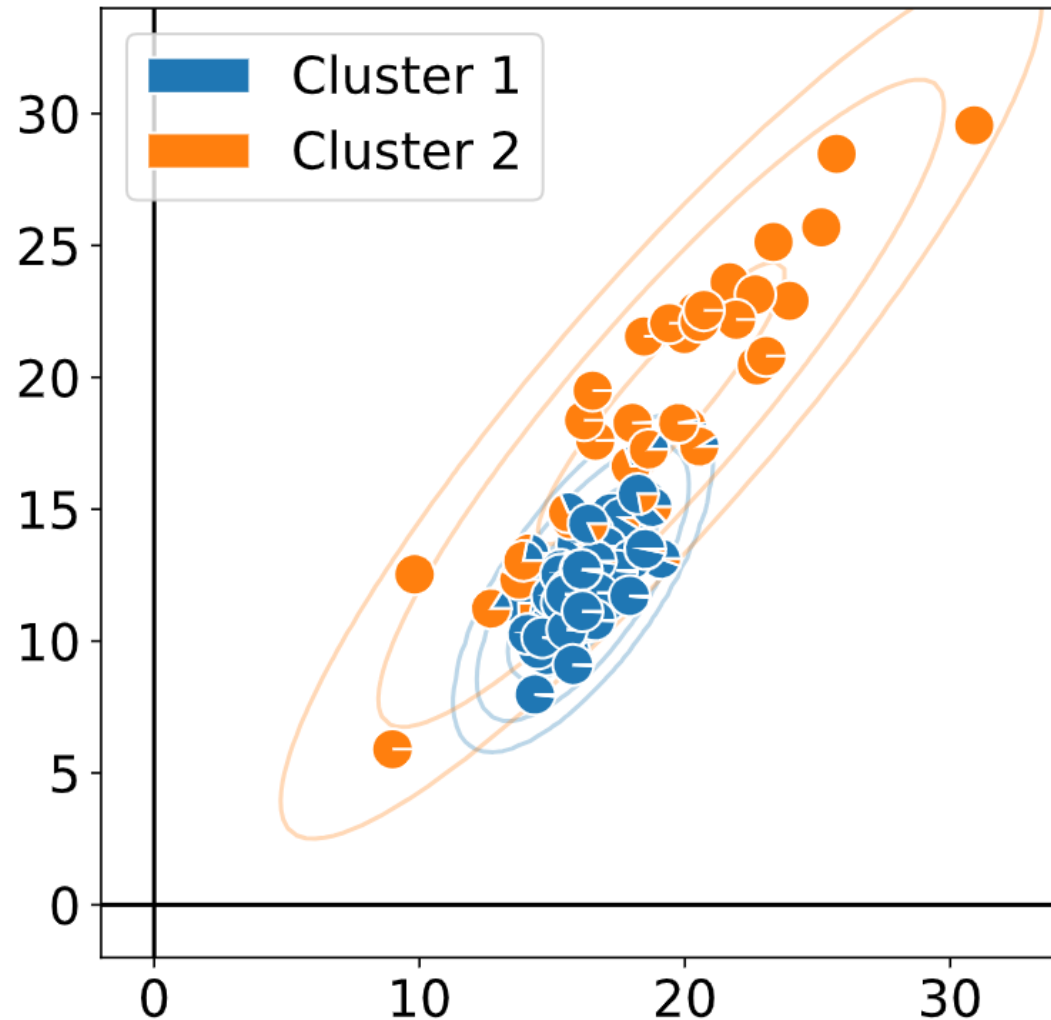
$$p(x_2 \mid cat, \mu_{cat,2}, \sigma_{cat,2})$$

$$p(x_1 \mid dog, \mu_{dog,1}, \sigma_{dog,1})$$

$$p(x_2 \mid dog, \mu_{dog,2}, \sigma_{dog,2})$$

C) None of the above

Sampling Cats and Dogs



B) Bernoulli + Gaussian

$$p(Y = cat \mid \phi)$$

$$p(x_1 \mid cat, \mu_{cat,1}, \sigma_{cat,1})$$

$$p(x_2 \mid cat, \mu_{cat,2}, \sigma_{cat,2})$$

$$p(x_1 \mid dog, \mu_{dog,1}, \sigma_{dog,1})$$

$$p(x_2 \mid dog, \mu_{dog,2}, \sigma_{dog,2})$$

Generative Stories

Generative vs Discriminative Modeling

Discriminative: $p(y | x)$

Generative: $p(y | x) = \alpha p(x, y) = \alpha p(x | y) p(y)$

Model assumptions vs Data

- Discriminative:
- Generative:

Generative Story Examples

Spam Generation

Discriminative

Generative

Generative + Naive Bayes