

CS 307

SPRING 2024

DALPIAZ

GENERATIVE
MODELS

DISCRIMINATIVE MODELS

DIRECTLY MODEL

$$P[Y = k \mid X = x]$$

- KNN
- TREE
- LOGISTIC REGRESSION

GIVEN MODEL, COULD ONLY GENERATE NEW Y DATA GIVEN X.

GENERATIVE MODELS

- MODEL FULL JOINT DISTRIBUTION

$$P[Y = k, X = x]$$

(Note: A red arrow points from the word "AND" written above the comma to the comma itself.)

- GIVEN MODEL, COULD GENERATE NEW X AND Y DATA!

CLASSIFICATION WITH GENERATIVE MODELS

$$P[Y=k | X=x] = \frac{P[Y=k, X=x]}{P[X=x]}$$

"AND" ↓

How TO MODEL? ↓

P[Y=k]
AND P[X=x | Y=k] } BASES

WE KNOW WHAT TO DO FROM HERE.

BAYES THEOREM

"GIVEN" ↓

"CONDITIONS ON B" ↗

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)}$$

MULTIPLICATION RULE ↖

"FLIPPED CONDITIONAL" ↘

DEFINITION OF CONDITIONAL PROBABILITY ↗

WHAT IF WE DON'T KNOW THIS? ↖

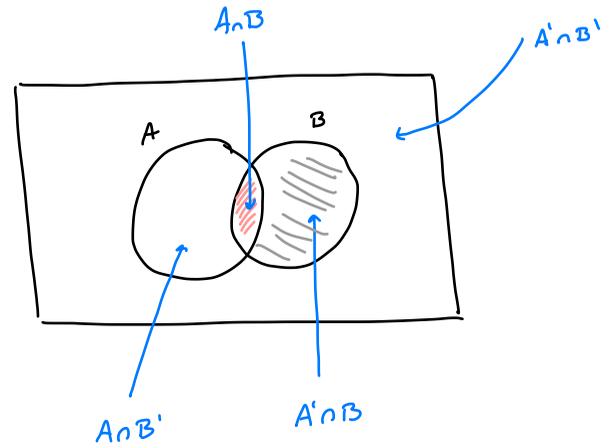
The diagram shows the derivation of Bayes' Theorem. It starts with the conditional probability $P(A|B)$ on the left. A red arrow labeled "GIVEN" points to it from above. A blue arrow labeled "CONDITIONS ON B" points to it from below. This is followed by an equals sign. A blue arrow labeled "DEFINITION OF CONDITIONAL PROBABILITY" points to this equals sign from below. The next part of the equation is the fraction $\frac{P(A \cap B)}{P(B)}$. A blue arrow labeled "MULTIPLICATION RULE" points from this fraction to the next fraction, $\frac{P(A)P(B|A)}{P(B)}$. A red arrow labeled "FLIPPED CONDITIONAL" points to $P(B|A)$ in the numerator of the second fraction from above. A grey arrow labeled "WHAT IF WE DON'T KNOW THIS?" points to the denominator $P(B)$ of the second fraction from below.

LAW OF TOTAL PROBABILITY

LOTP

$$P(B) = ?$$

FOR B TO OCCUR, EITHER A OR A' MUST OCCUR!



MULTIPLICATION RULE

$$P(B) = P(A \cap B) + P(A' \cap B) = P(A)P(B|A) + P(A')P(B|A')$$

A OCCURS A' OCCURS THEN B



BAYES THEOREM, REWRITTEN

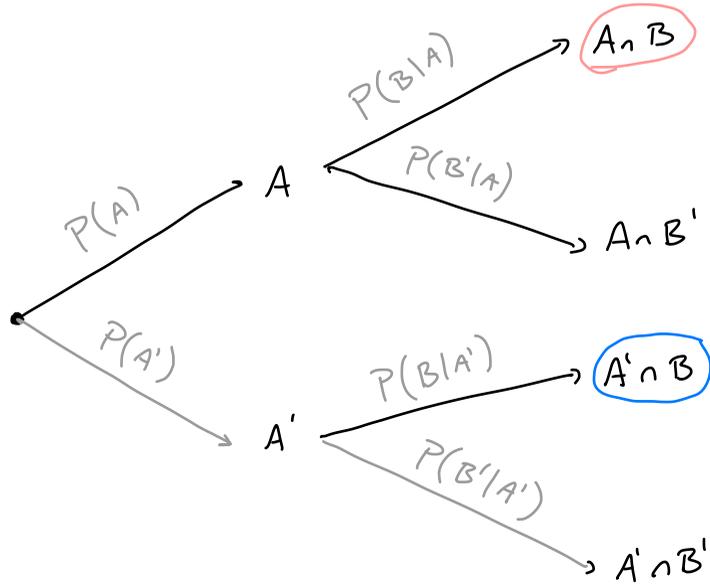
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)}$$

$$= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$$

LAW OF TOTAL
PROBABILITY



"TREE VIEW"



$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$= \frac{\underline{P(A \cap B)}}{\underline{P(A \cap B)} + \underline{P(A' \cap B)}}$$

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

$$P(A' \cap B) = P(A')P(B|A')$$

BAYES FOR CATEGORICAL Y , CONTINUOUS X

$$P(Y=A) = \pi_A$$

$$P(Y=B) = \pi_B$$

NOTATION

$$f_{X|Y=A}(x) = f_A(x)$$

$$f_{X|Y=B}(x) = f_B(x)$$

NOTATION

PDF OF X
WHEN $Y=A$

PDF ON X
WHEN $Y=B$

$$P(Y=A|X=x) = \frac{\pi_A f_A(x)}{\pi_A f_A(x) + \pi_B f_B(x)}$$

$$P(Y=B|X=x) = ???$$

BAYES FOR CATEGORICAL Y, CONTINUOUS X

$$P(Y=A) = \pi_A$$

$$f_{X|Y=A}(x) = f_A(x)$$

$$P(Y=B) = \pi_B$$

$$f_{X|Y=B}(x) = f_B(x)$$

$$P(Y=C) = \pi_C$$

$$f_{X|Y=C}(x) = f_C(x)$$

$$P(Y=A|X=x) = \frac{\pi_A f_A(x)}{\pi_A f_A(x) + \pi_B f_B(x) + \pi_C f_C(x)}$$

LOTP BY THREE POSSIBILITIES FOR Y



BAYES FOR CATEGORICAL Y, CONTINUOUS X

$$P(Y=k) = \pi_k$$

NUMBER OF
Y CATEGORIES

$$\sum_{g=1}^G \pi_g = 1$$

PRIOR PROBABILITIES
BEFORE SEEING DATA

$$f_{X|Y=k}(x) = f_k(x)$$

LIKELIHOODS
OF DATA

$$P_k(x) = P(Y=k | X=x) = \frac{\pi_k f_k(x)}{\sum_{g=1}^G \pi_g f_g(x)}$$

POSTERIOR PROBABILITY
UPDATED AFTER SEEING DATA

EXAMPLE

$$x = 3.4$$

PRIORS

$$\pi_A = 0.20$$

$$\pi_B = 0.50$$

$$\pi_C = 0.30$$

LIKELIHOODS

$$X | Y = A \sim \mathcal{N}(\mu = 2, \sigma = 1)$$

$$X | Y = B \sim \mathcal{N}(\mu = 3, \sigma = 2)$$

$$X | Y = C \sim \mathcal{N}(\mu = 4, \sigma = 1)$$

$$f_A(3.4) = ?$$

$$f_B(3.4) = ?$$

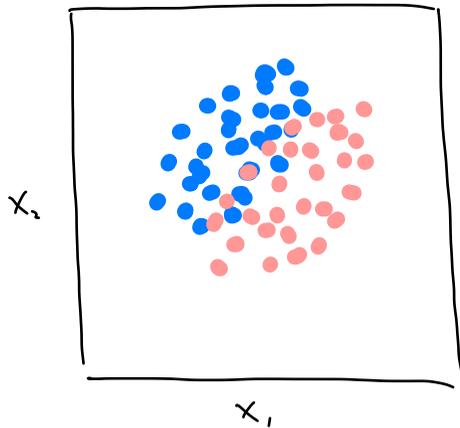
$$f_C(3.4) = ?$$

POSTERIOR

$$P(Y = C | X = 3.4) = \frac{\pi_C f_C(3.4)}{\pi_A f_A(3.4) + \pi_B f_B(3.4) + \pi_C f_C(3.4)} =$$

SEE PYTHON!

GENERATIVE SETUP



pdf $f_1(x)$

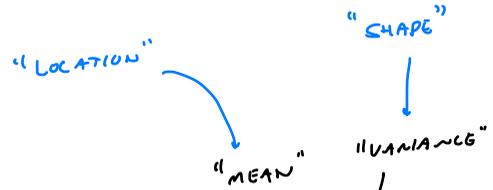


$$(x_1, x_2) \mid Y=1 \sim \text{MVN}(\mu_1, \Sigma_1)$$



pdf $f_2(x)$

$$(x_1, x_2) \mid Y=0 \sim \text{MVN}(\mu_0, \Sigma_0)$$



$$P[Y=1] = \pi_1$$

$$P[Y=0] = \pi_0$$

COVARIANCE

X_1, X_2

Σ

=

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

Var(X_1)

Cov(X_1, X_2)

Var(X_2)

$$\sigma_{12} = \sigma_{21}$$

COVARIANCE
MATRIX

CORR(X_1, X_2)

↓

$$\sigma_{12} = \rho_{12} \sigma_1 \sigma_2$$

Similar in higher
dimensions

POSTERION

BAYES THEOREM

$$P[Y=1 | X=x] = \frac{\pi_1 f_1(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}$$

MUM

π_0, π_1 "PRIOR" PROBABILITIES

$f_1(x), f_2(x)$ LIKELIHOODS

OR SET DIRECTLY IF KNOWN
OR ASSUMED.

NEED TO ESTIMATE

π_0, π_1

μ_1, μ_2

Σ_1, Σ_2

How?

MLE PROBABLY

THREE WAYS TO MODEL $f_k(x)$

→ LINEAR
LDA

$$\Sigma = \Sigma_1 = \Sigma_2 = \dots = \Sigma_C$$

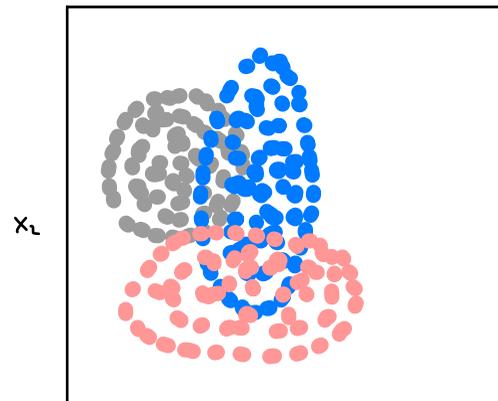
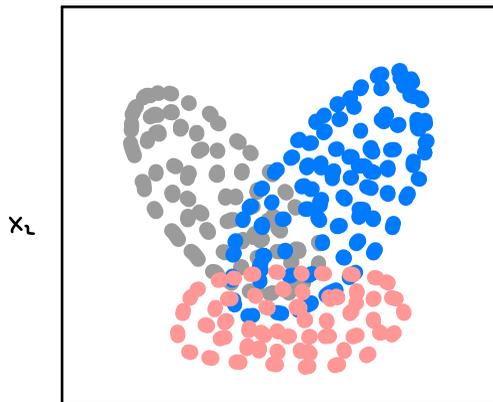
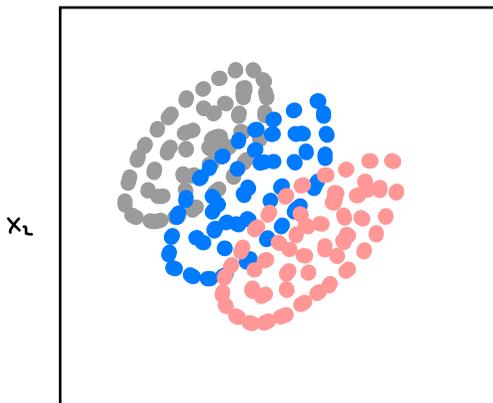
→ QUADRATIC
GDA

$$\Sigma_K$$

→ NAIVE BAYES

↓
NB

$$\Sigma_K = \begin{bmatrix} \sigma_{k1}^2 & & 0 \\ & \dots & \\ 0 & & \sigma_{kp}^2 \end{bmatrix}$$



NAIVE BAYES

NAIVE \Rightarrow GIVEN Y, X_1, \dots, X_p IND
ASSUMPTION

PRODUCT OF UNIVARIATE NORMALS
PRODUCT OF PDFs!

$$f_k(x_1, x_2, \dots, x_p) = \prod_{j=1}^p f_{kj}(x_j)$$

MMN

pdf OF FEATURE j GIVEN $Y=k$

$$f_{kj}(x_j) = f_{x_j|Y=k}(x_j) \sim \mathcal{N}(\mu_{kj}, \sigma_{kj}^2)$$

NEED TO ESTIMATE

NO NEED TO ESTIMATE COVARIANCES!!!

ESTIMATION IN NAIVE BAYES

$$n_k = \sum_{i=1}^n I(y_i = k) \quad \leftarrow \text{\# TIMES } Y_i = k \text{ IN DATA}$$

$$\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^n I(y_i = k) \quad \leftarrow \text{PROPORTION OF } Y_i = k \text{ IN DATA}$$

$$\hat{\mu}_{kj} = \frac{1}{n_k} \sum_{i=1}^n x_j \cdot I(y_i = k) \quad \leftarrow \text{MEAN OF } X_j \text{ WHEN } y_i = k$$

$$\hat{\sigma}_{kj} = \sqrt{\frac{1}{n_k} \sum_{i=1}^n (x_j \cdot I(y_i = k) - \hat{\mu}_{kj})^2} \quad \leftarrow \text{SD OF } X_j \text{ WHEN } y_i = k$$

INDICATOR FUNCTION \rightarrow

$$I(y_i = k) = \begin{cases} 1 & \text{IF } y_i = k \\ 0 & \text{IF } y_i \neq k \end{cases}$$

In PYTHON

sklearn

Linear Discriminant Analysis
Quadratic Discriminant Analysis



priors_
means_
covariance_



ESTIMATES

Gaussian NB



class_prior_
theta_
var_



ESTIMATES