CS 307

Fall 2023

Dalpiaz

Week 10

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

"and"

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

- Given model, could generate new X and Y data!

Classification with Generative Models

"and"

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

How to model?

$P[Y=k]$

and $P[X=x \mid Y=k]$ $\Big\}$ Bayes

We know what to do from here.

Bayes Theorem

"Given"

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

Multiplication rule

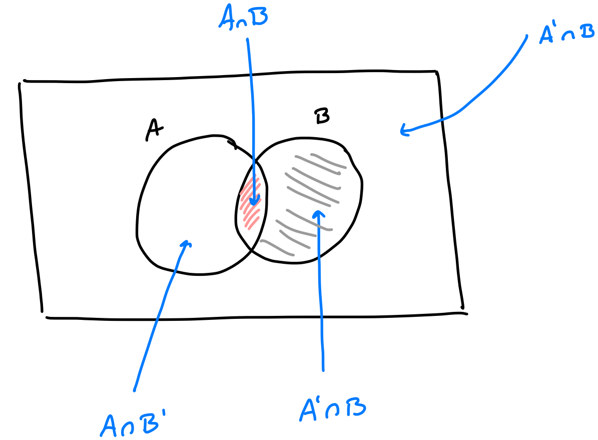
"Flipped conditional"

"Conditioned on B"

Definition of conditional probability

What if we don't know this?

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

then B

A' occurs

then B

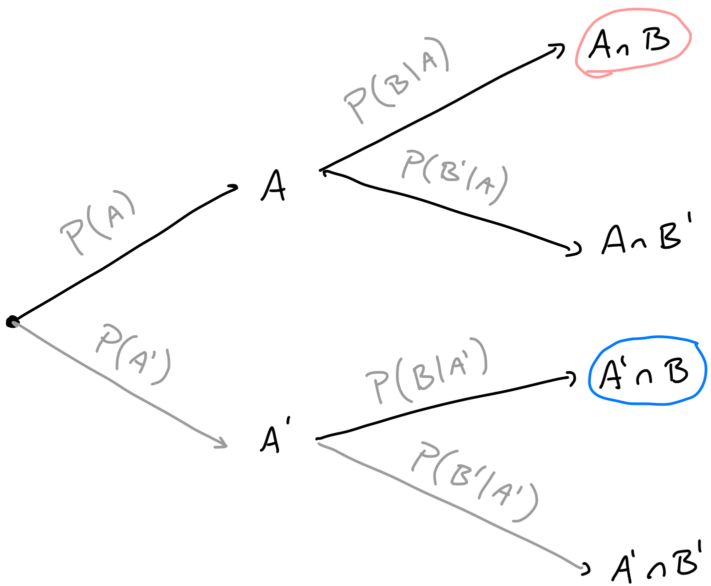
Bayes Theorem, Rewritten

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

Law of total probability

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

"TREE VIEW"



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

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

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

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

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

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

$$P(Y=B \mid X=x)= ???$$

Bayes for categorical Y, continuous X

$P(Y=A) = \pi_A$

$P(Y=B) = \pi_B$

$P(Y=C) = \pi_C$

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

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

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

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

Loop by three possibilities for Y

BAYES FOR CATEGORICAL Y, CONTINUOUS X

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

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

NUMBER OF
Y CATEGORIES

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

PRIOR PROBABILITIES
BEFORE SEEING DATA

LIKELIHOODS
OF DATA

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

POSTERIOR PROBABILITY
UPDATED AFTER SEEING DATA

EXAMPLE

$x = 3.4$

PRIORS

$\pi_A = 0.20$

$\pi_B = 0.50$

$\pi_C = 0.30$

LIKELIHOODS

$X \mid Y = A \sim N(\mu = 2, \sigma = 1)$

$X \mid Y = B \sim N(\mu = 3, \sigma = 2)$

$X \mid Y = C \sim N(\mu = 4, \sigma = 1)$

$f_A(3.4) = ?$

$f_B(3.4) = ?$

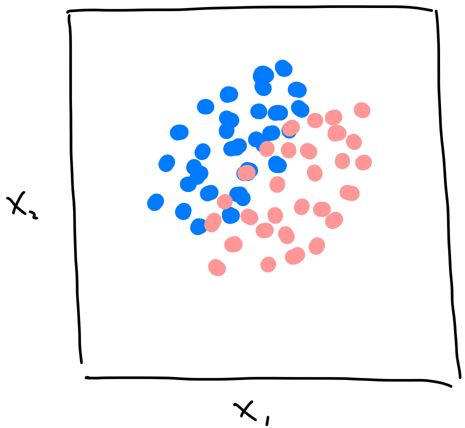
$f_C(3.4) = ?$

POSTERIOR

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

SEE PYTHON!

Generative Setup



"location" "shape"

"mean" "variance"

pdf $f_1(x)$

$$(X_1, X_2) \mid Y = 1 \sim MVN(\mu_1, \Sigma_1)$$

$$(X_1, X_2) \mid Y = 0 \sim MVN(\mu_0, \Sigma_0)$$

pdf $f_2(x)$

$$P[Y=1] = \pi_1 \qquad P[Y=0] = \pi_0$$

Covariance $\quad X_1, X_2$

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

Σ: Covariance matrix

σ_1^2: Var(X_1)

σ_{12}: Cov(X_1, X_2)

σ_2^2: Var(X_2)

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

$$\sigma_{12} = \underbrace{\rho_{12}}_{\text{Corr}(X_1, X_2)} \sigma_1 \sigma_2$$

Similar in higher dimensions

BAYES THEOREM

MVN

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

POSTERIOR

π_0, π_1 "PRIOR" PROBABILITIES

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

OR SET DIRECTLY IF KNOWN
OR ASSUMED.

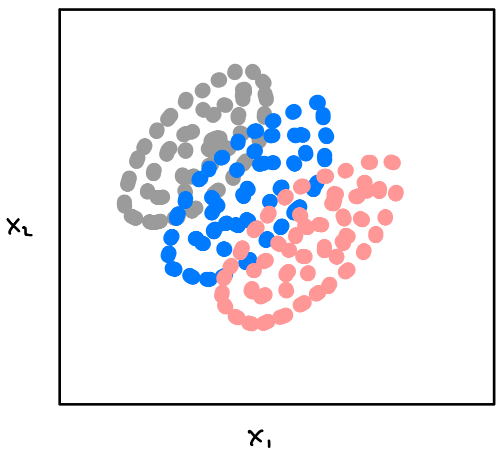
NEED TO ESTIMATE π_0, π_1, μ_1, μ_2, Σ_1, Σ_2 How? MLE PROBABLY

Three ways to model $f_k(x)$

LDA

LINEAR

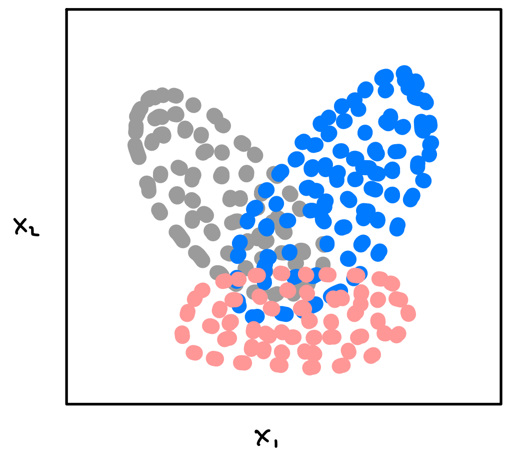
$$\Sigma = \Sigma_1 = \Sigma_2 = \dots = \Sigma_G$$



QDA

QUADRATIC

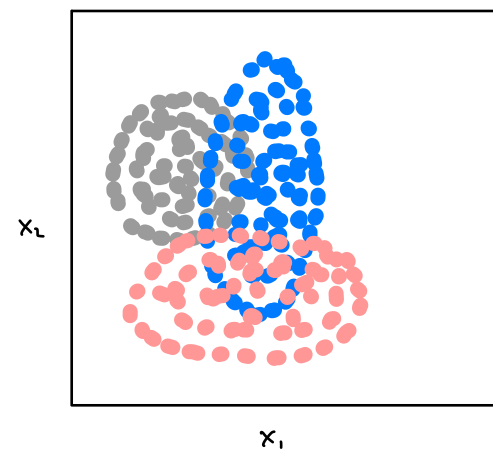
$$\Sigma_k$$



NB

NAIVE BAYES

$$\Sigma_k = \begin{bmatrix} \sigma_{k1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{kp}^2 \end{bmatrix}$$



NAIVE BAYES

NAIVE $\Rightarrow$ GIVEN Y, $X_1, \ldots X_p$ IND

ASSUMPTION

PRODUCT OF UNIVARIATE NORMALS

PRODUCT OF PDFs!

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

MVN

pdf OF FEATURE j GIVEN $Y=k$

NEED TO ESTIMATE

$$f_{kj}(x_j) = f_{x_j \mid Y=k}(x_j) \sim N(\mu_{kj}, \sigma_{kj})$$

$Y=k$

x_j

NO NEED TO ESTIMATE COVARIANCES!!!

Estimation in Naive Bayes

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

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

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

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

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

In Python

sklearn

Linear Discriminant Analysis → priors_
Quadratic Discriminant Analysis

$\left.\begin{array}{l}\text{priors\_}\\ \text{means\_}\\ \text{covariance\_}\end{array}\right\}$ ESTIMATES

Gaussian NB → $\left.\begin{array}{l}\text{class\_prior\_}\\ \text{theta\_}\\ \text{var\_}\end{array}\right\}$ ESTIMATES