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

"AND"

$$P\left[Y = k, X = x\right]$$

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

Classification with Generative Models

"and"

How to model?

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

$P[Y=k]$

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

We know what to do from here.

Bayes Theorem

"FLIPPED CONDITIONAL"

MULTIPLICATION RULE

"GIVEN"

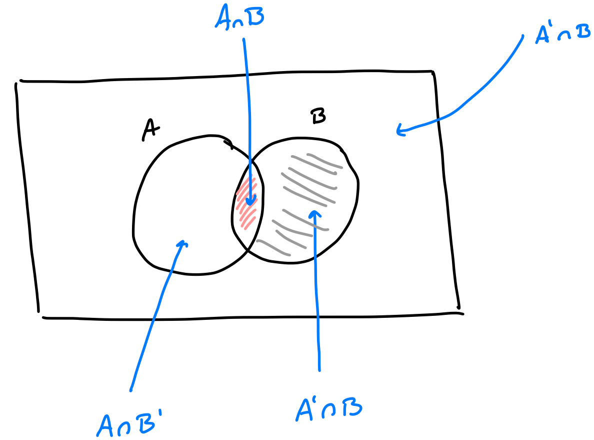
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\,P(B|A)}{P(B)}$$

"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

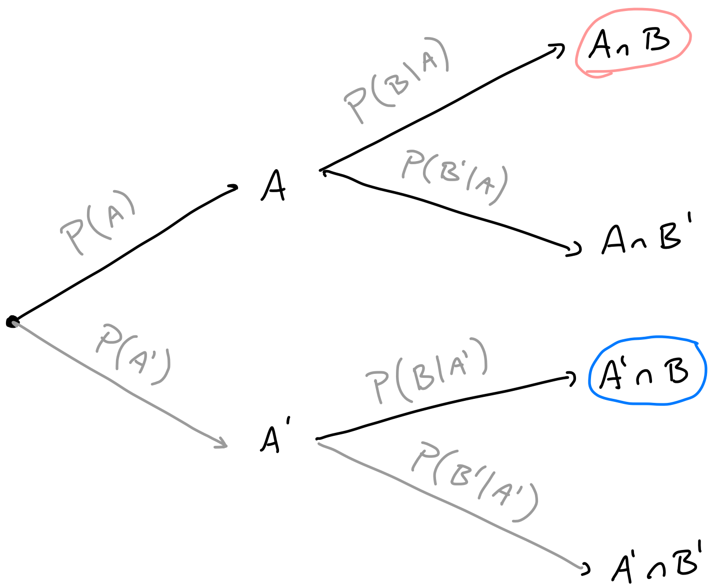
Bayes Theorem, Rewritten

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

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

$$P(Y=A) = \pi_A \qquad f_{X|Y=A}(x) = f_A(x)$$

$$P(Y=B) = \pi_B \qquad f_{X|Y=B}(x) = f_B(x)$$

$$P(Y=C) = \pi_C \qquad f_{X|Y=C}(x) = f_C(x)$$

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

LOOP BY THREE POSSIBILITIES FOR Y

BAYES FOR CATEGORICAL Y, CONTINUOUS X

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

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

NUMBER OF Y CATEGORIES

PRIOR PROBABILITIES BEFORE SEEING DATA

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

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

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

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

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

$f_A(3.4) = ?$

$f_B(3.4) = ?$

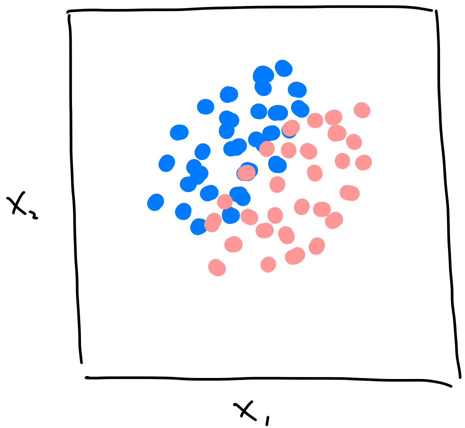
$f_C(3.4) = ?$

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" → "mean"

"Shape" → "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 $f_2(x)$

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

COVARIANCE

X_1, X_2

VAR(X_1)

COV(X_1, X_2)

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

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

VAR(X_2)

COVARIANCE MATRIX

CORR(X_1, X_2)

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

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 $f_k(x)$

LINEAR

LDA

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

QUADRATIC

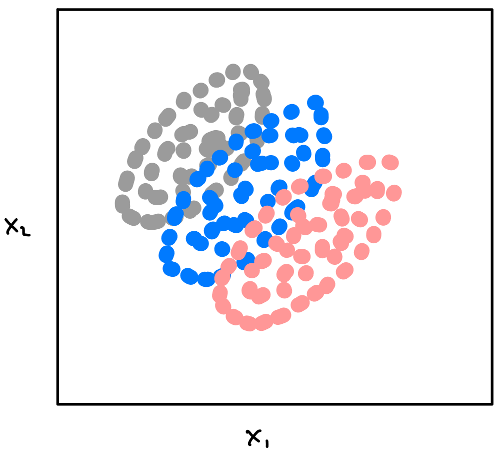
QDA

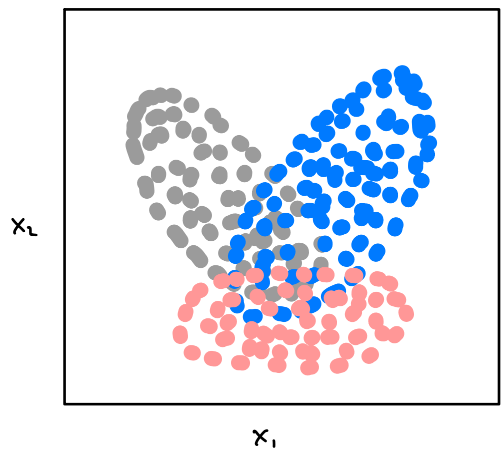
$$\Sigma_k$$

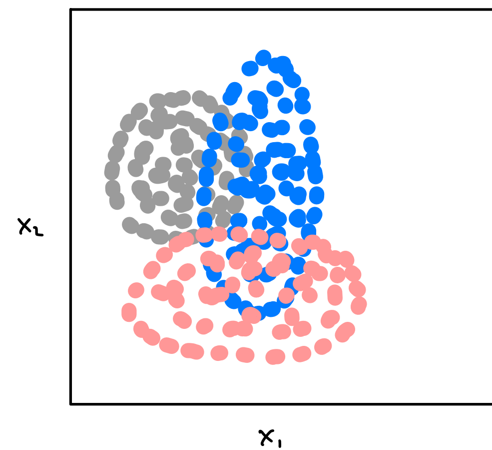
NAIVE BAYES

NB

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







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

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

$Y=k$

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
priors_
means_
covariance_
} ESTIMATES

Gaussian NB → class_prior_
theta_
var_
} ESTIMATES