CS 307

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"

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"

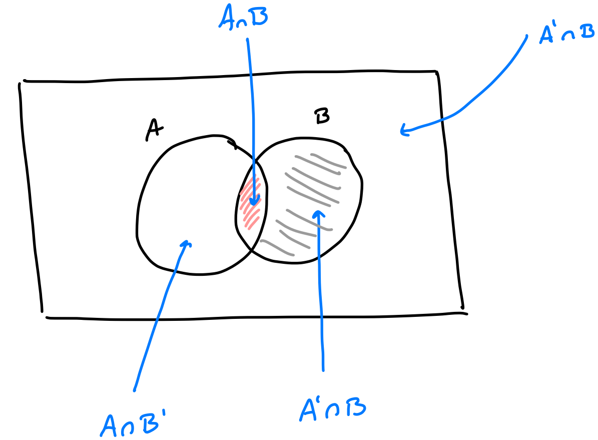
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\,P(B|A)}{P(B)}$$

"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!

$$P(B) = P(A \cap B) + P(A' \cap B) \overset{\text{MULTIPLICATION RULE}}{=} \underbrace{P(A)}_{\text{A occurs}}\,\underbrace{P(B|A)}_{\text{then B}} + \underbrace{P(A')}_{\text{A' occurs}}\,\underbrace{P(B|A')}_{\text{then B}}$$

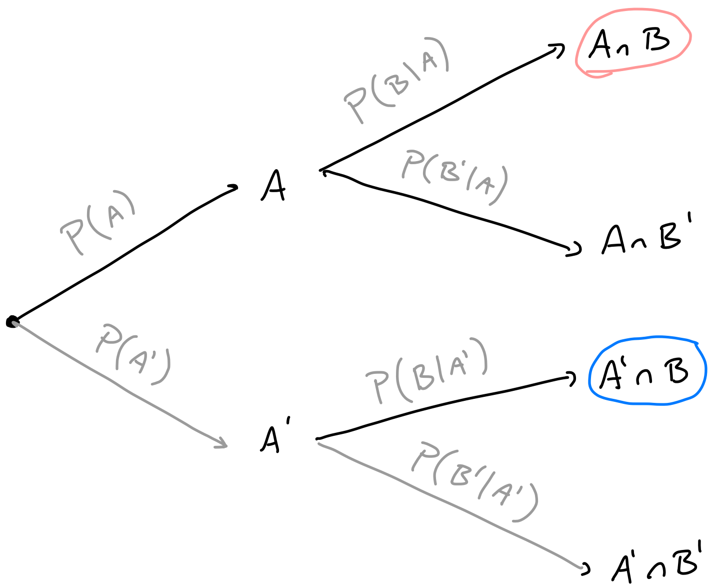
Bayes Theorem, Rewritten

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

Law of Total Probability

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

"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

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

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

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

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

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

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

$f_A(3.4) = ?$

$f_B(3.4) = ?$

$f_C(3.4) = ?$

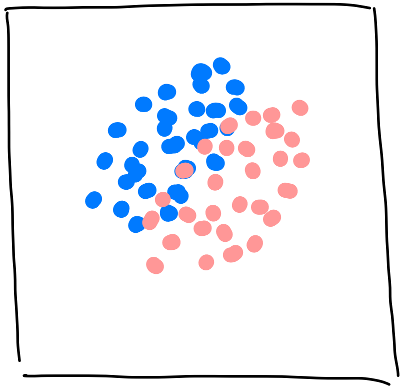
Posterior

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

SEE PYTHON!

Generative Setup

X_2



X_1

"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 $f_2(x)$

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

Covariance $\quad X_1, X_2$

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

Var(X_1)

Cov(X_1, X_2)

Var(X_2)

Covariance matrix

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

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

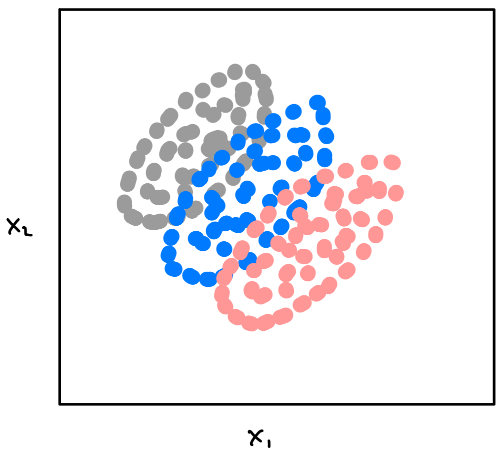
OR SET DIRECTLY IF KNOWN
OR ASSUMED.

NEED TO ESTIMATE $\left.\begin{array}{l}\pi_0, \pi_1 \\ \mu_1, \mu_2 \\ \Sigma_1, \Sigma_2\end{array}\right\}$ How? MLE PROBABLY

THREE WAYS TO MODEL $f_k(x)$

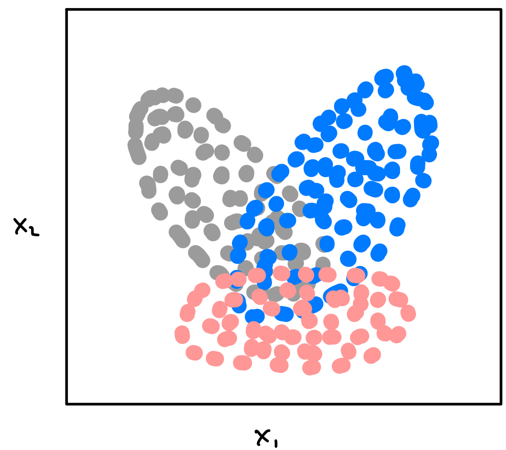
LDA → LINEAR

$$\Sigma = \Sigma_1 = \Sigma_2 = \cdots = \Sigma_G$$



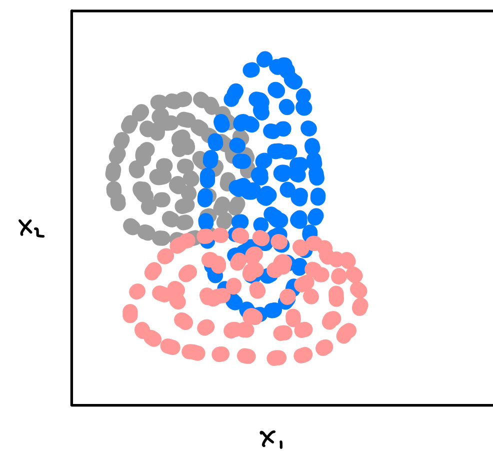
QDA → QUADRATIC

$$\Sigma_k$$



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

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

$Y=k$

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

GaussianNB → class_prior_

class_prior_
theta_
var_

ESTIMATES