IIT KHARAGPUR

NPTEL ONLINE
CERTIFICATION COURSES

Data Mining

Week 3:Bayes Classification

Pabitra Mitra

Computer Science and Engineering, IIT Kharagpur

NPTEL

Data Mining

Bayes Classification

Pabitra Mitra

Computer Science and Engineering, IIT Kharagpur

NPTEL

A Simple Species Classification Problem

- Measure the *length* of a fish, and decide its class
  - Hilsa or Tuna

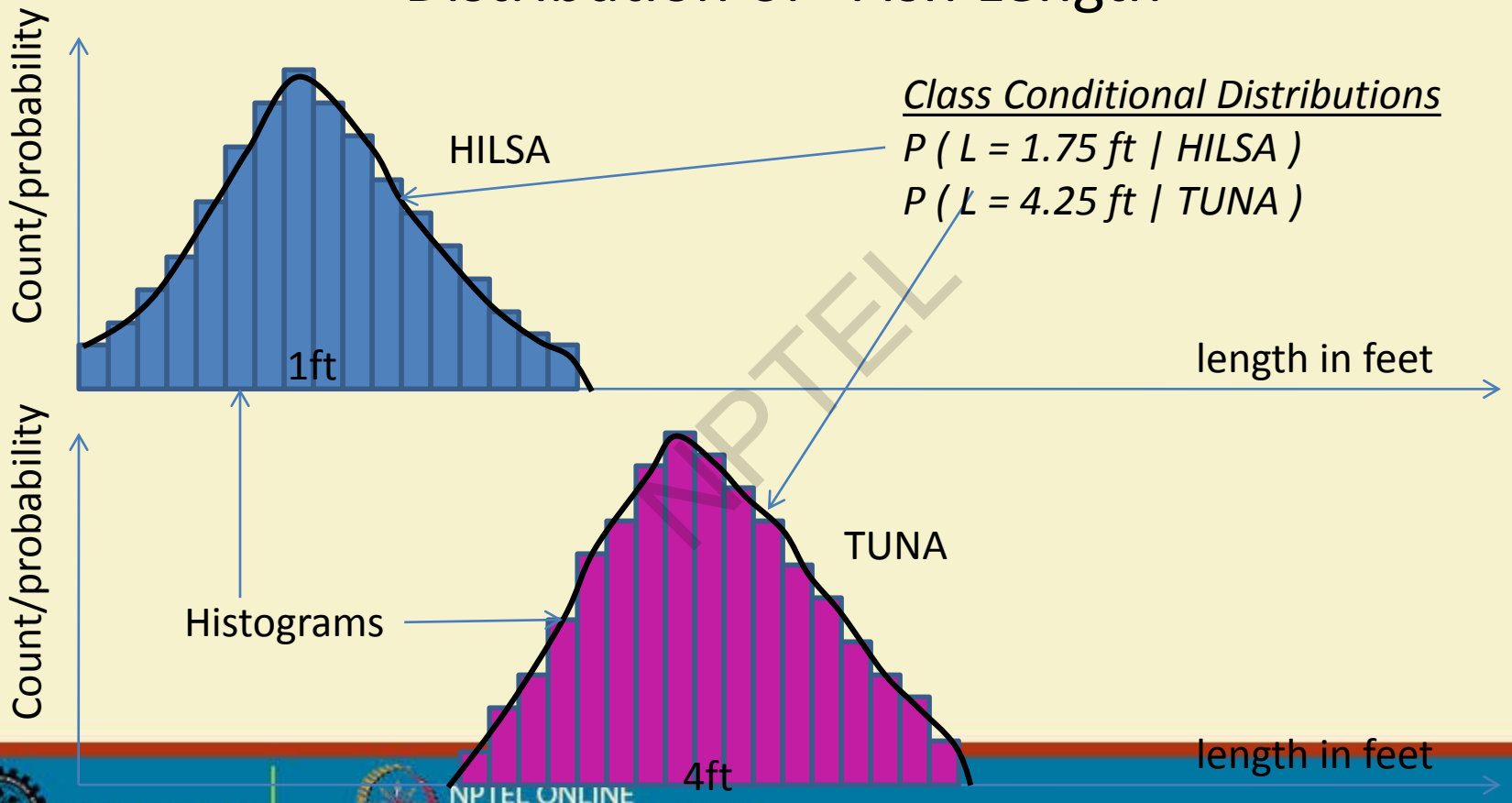
NPTEL

Collect Statistics …

Population for Class Hilsa

Population for Class Tuna

Distribution of “Fish Length”



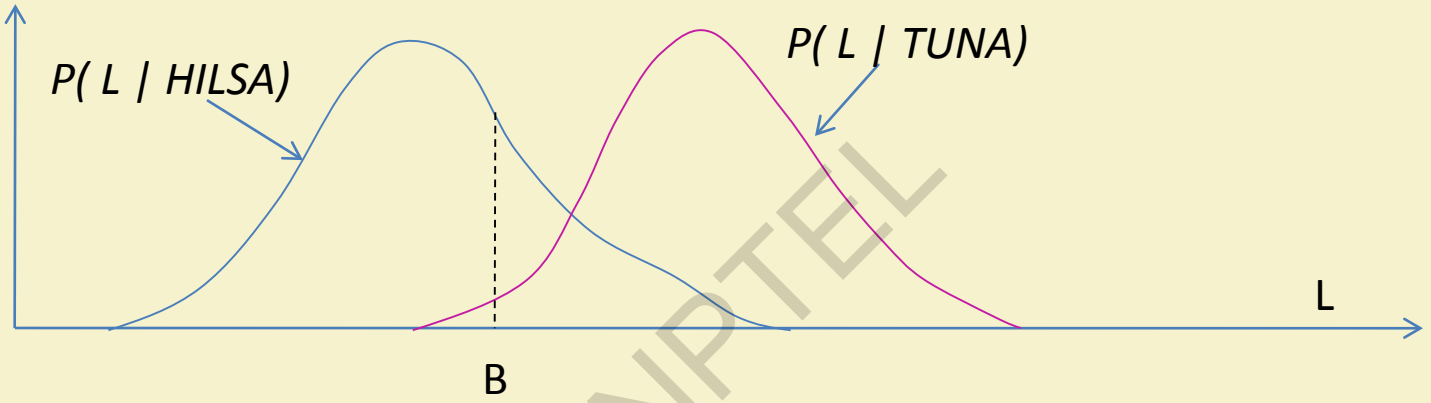
NPTEL

Decision Rule

- If length L ≤ B
  - HILSA
- ELSE
  - TUNA
- What should be the value of B (“boundary” length) ?
  - Based on population statistics

NPTEL

Error of Decision Rule



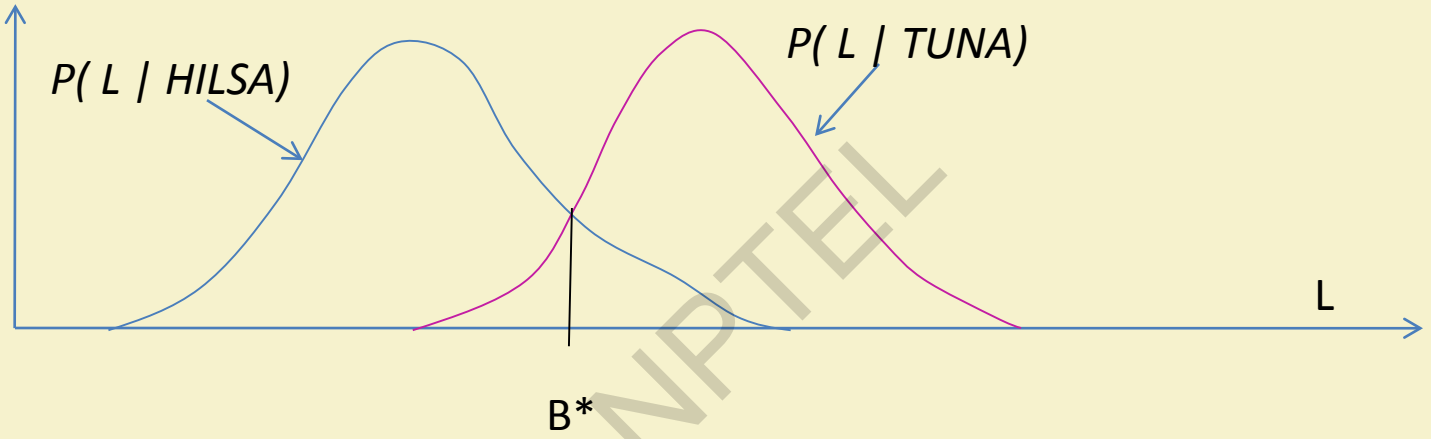
NPTEL

Errors: Type 1 + Type 2,

Type 1: Actually Tuna, Classified as Hilsa (area under pink curve to the left of a B)

Type 2: Actually Hilsa, Classified as Tuna (area under blue curve to the right of a B)

Optimal Decision Rule



B*: Optimal Value of B, (Optimal Decision Boundary)

Minimum Possible Error

$$P(B^* \mid HILSA) = P(B^* \mid TUNA)$$

If Type 1 and Type 2 errors have different costs : optimal boundary shifts

NPTEL

Species Identification Problem

- Measure lengths of a (sizeable) population of Hilsa and Tuna fishes
- Estimate Class Conditional Distributions for Hilsa and Tuna classes respectively
- Find Optimal Decision Boundary B* from the distributions
- Apply Decision Rule to classify a newly caught (and measured) fish as either Hilsa or Tuna
  - (with minimum error probability)

NPTEL

Location/Time of Experiment

- Calcutta in Monsoon
  - More Hilsa few Tuna
- California in Winter
  - More Tuna less Hilsa
- Even a 2ft fish is likely to be Hilsa in Calcutta
- a 1.5ft fish may be Tuna in California

NPTEL

Apriori Probability

- Without measuring length what can we guess about the class of a fish
  - Depends on location/time of experiment
    - Calcutta : Hilsa, California: Tuna
- Apriori probability: *P(HILSA), P(TUNA)*
  - Property of the frequency of classes during experiment
    - Not a property of length of the fish
  - Calcutta: *P(Hilsa) = 0.90, P(Tuna) = 0.10*
  - California: *P(Tuna) = 0.95, P(Hilsa) = 0.05*
  - London*: P(Tuna) = 0.50, P(Hilsa) = 0.50*
- Also a determining factor in class decision along with class conditional probability

NPTEL

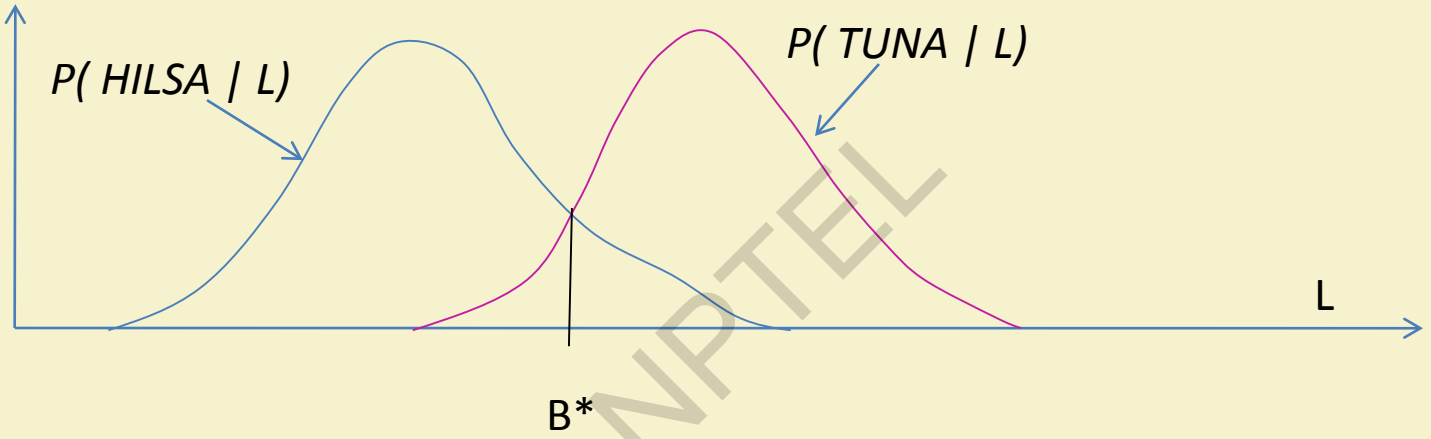
Classification Decision

- We consider the product of *Apriori* and *Class conditional* probability factors
- *Posteriori probability (Bayes rule)*
  - *P(HILSA | L = 2ft) = P(HILSA) x P(L=2ft | HILSA) / P(L=2ft)*
  - *Posteriori ≈ Apriori x Class conditional*
  - *denominator is constant for all classes*
- Apriori: Without any measurement - based on just location/time – what can we guess about class membership (estimated frm size of class populations)
- Class conditional: Given the fish belongs to a particular class what is the probability that its length is *L=2ft* (estimated from population)
- Posteriori: Given the measurement that the length of the fish is *L=2ft* what is the probability that the fish belongs to a particular class (obtained using Bayes rule from above two probabilities).
  - Useful in decision making using evidences/measurements.

NPTEL

Bayes Classification Rule (Bayes Classifier)

Posteriori Distributions



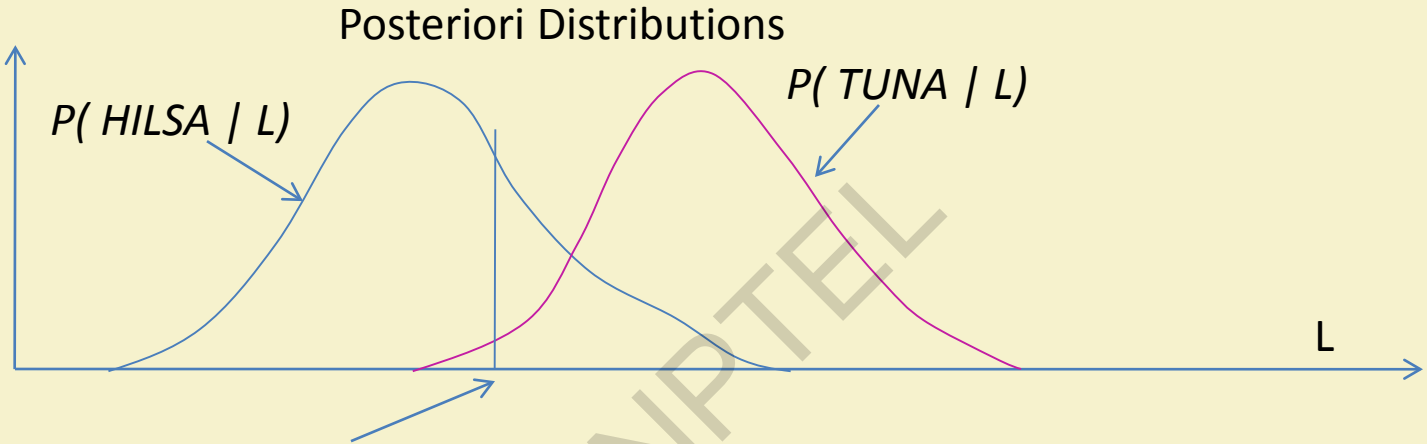
NPTEL

B*: Optimal Value of B, (Bayes Decision Boundary)

P (HILSA| L= B*) = P (TUNA | L = B*)

Minimum error probability: Bayes error

MAP Representation of Bayes Classifier

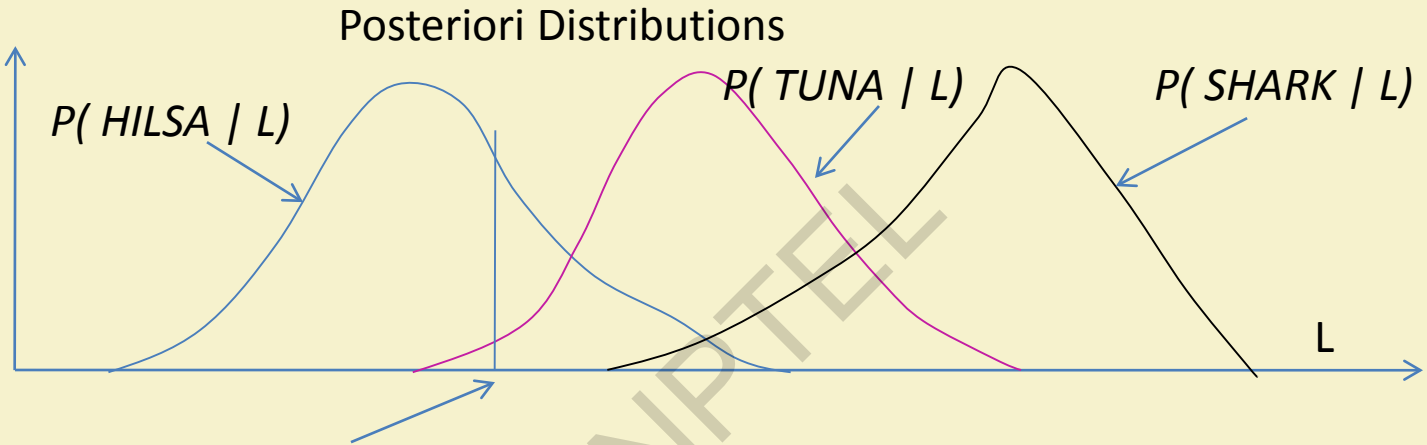


Hilsa has higher posteriori probability than Tuna for this length

Instead of finding decision boundary B*, state classification rule as:
Classify an object in to the class for which it has the highest posteriori prob.
(MAP: Maximum Aposteriori Probability)

NPTEL

MAP Multiclass Classifier

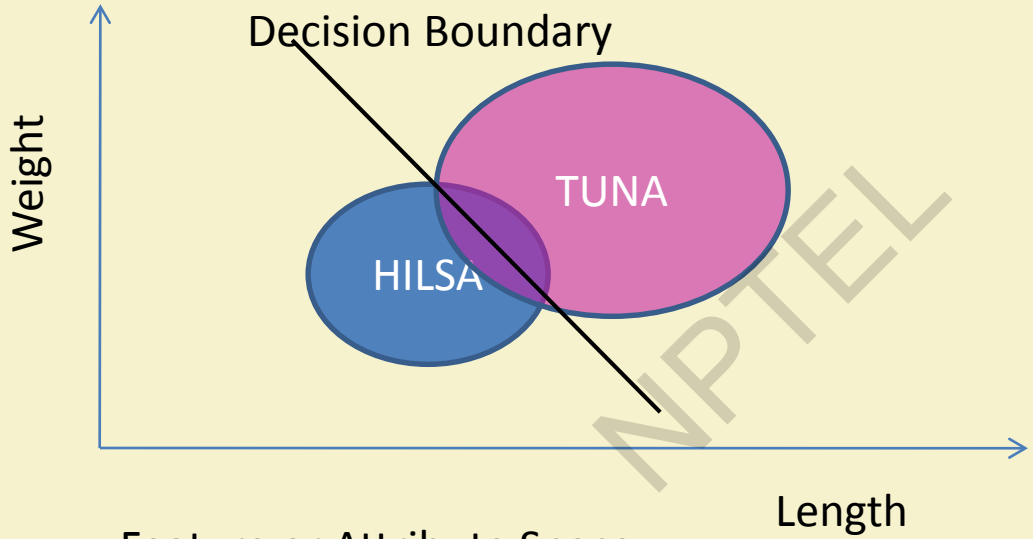


Hilsa has highest posteriori probability among all classes for this length

Classify an object in to the class for which it has the highest posteriori prob.
(MAP: Maximum Aposteriori Probability)

NPTEL

Multivariate Bayes Classifier

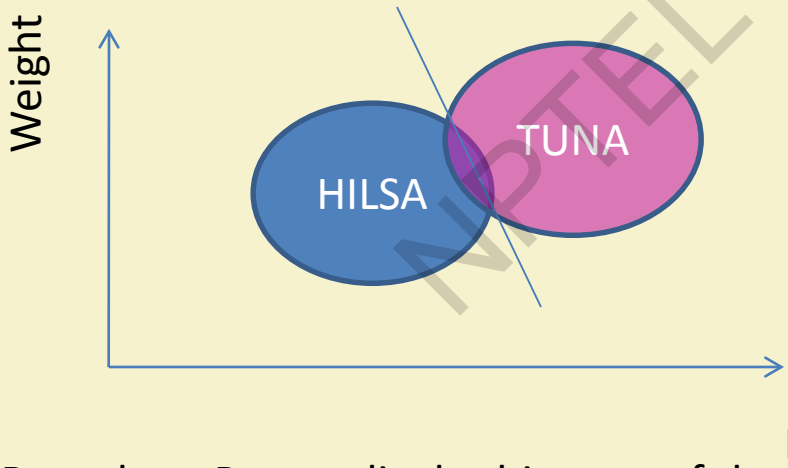


NPTEL

- Feature or Attribute Space
- Class Seperability

Decision Boundary: Normal Distribution

- Two spherical classes having different means, but same variance (diagonal covariance matrix with same variances)

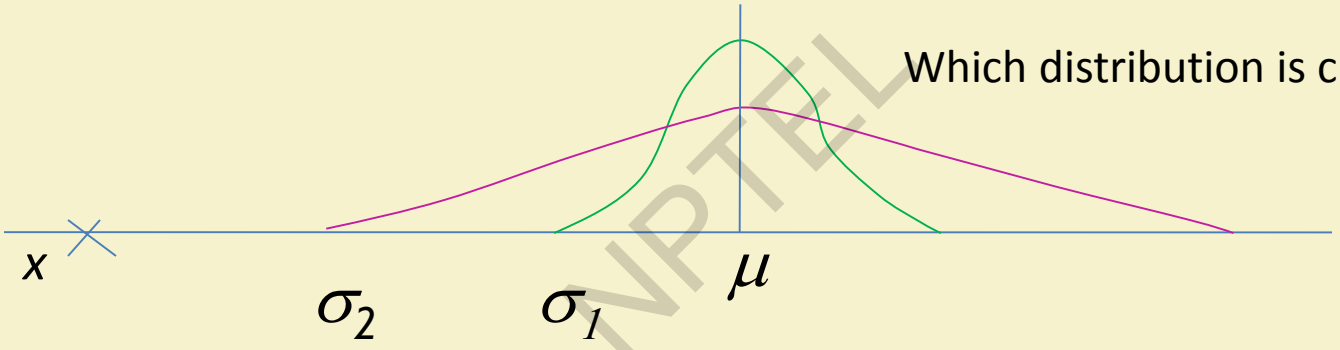


Decision Boundary: Perpendicular bisector of the mean vectors

NPTEL

Distances

- Two vectors: Euclidean, Minkowski etc
- A vector and a distribution: Mahalanobis, Bhattacharya



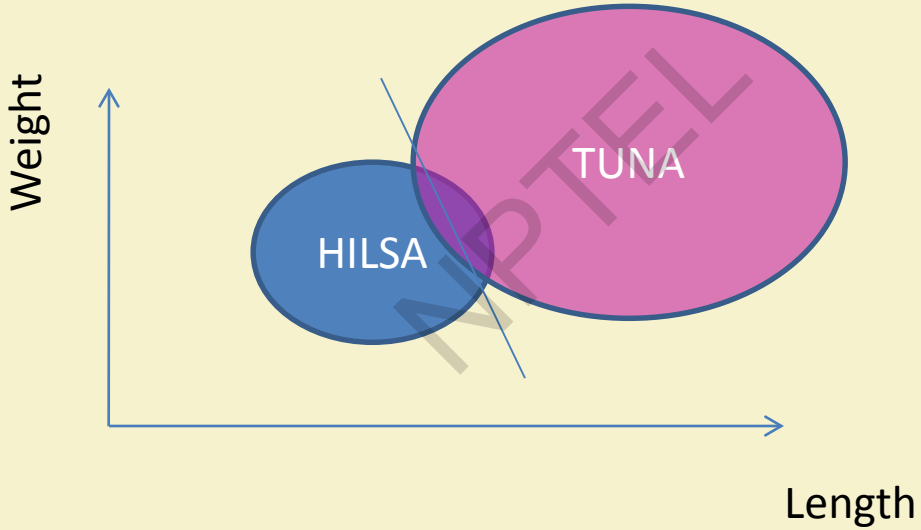
$$d_M = \frac{(x-\mu)^2}{\sigma}, d_M = (X-\mu)\Sigma^{-1}(X-\mu)^T$$

- Between two distributions: Kullback-Liebler Divergence

NPTEL

Decision Boundary: Normal Distribution

- Two spherical classes having different means and variances (diagonal covariance matrix with different variances)

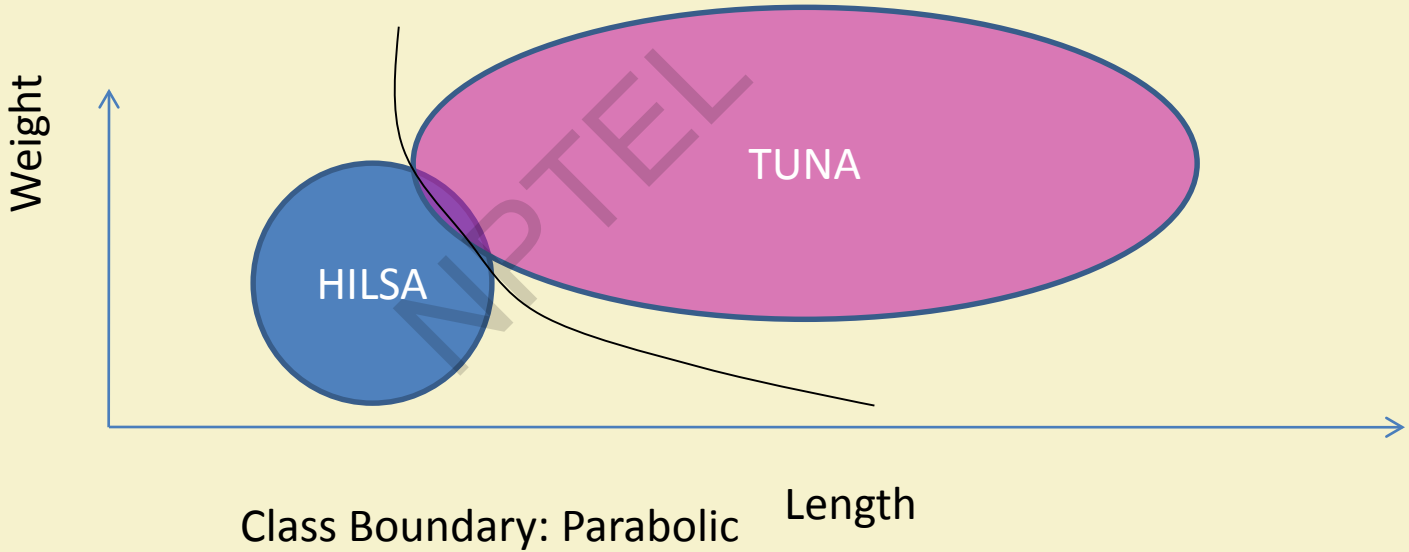


Boundary: Locus of equi-Mahalanobis distance points from the class distributions. (still a straight line)

NPTEL ONLINE CERTIFICATION COURSES

Decision Boundary: Normal Distribution

- Two elliptical classes having different means and variances (general covariance matrix with different variances)



NPTEL

Multivariate Bayesian Classifiers

- Approach:
  - compute the posterior probability P(C | A_1, A_2, …, A_n) for all values of C using the Bayes theorem

$$P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}$$

  - Choose value of C that maximizes
    P(C | A_1, A_2, …, A_n)
  - Equivalent to choosing value of C that maximizes
    P(A_1, A_2, …, A_n|C) P(C)
- How to estimate P(A_1, A_2, …, A_n | C)?

NPTEL

Example of Bayes Classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: attributes

M: mammals

N: non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

$P(A|M)P(M) > P(A|N)P(N)$

=> Mammals

NPTEL

Estimating Multivariate Class Distributions

- Sample size requirement
  - In a small sample: difficult to find a Hilsa fish whose length is 1.5ft and weight is 2 kilos, as compared to that of just finding a fish whose length is 1.5ft
  - *P(L=1.5, W=2 | Hilsa), P(L=1.5 | Hilsa)*
  - Curse of dimensionality
- Independence Assumption
  - Assume length and weight are independent
  - *P(L=1.5, W=2 | Hilsa) = P(L=1.5 | Hilsa) x P(W=2| Hilsa)*
  - Joint distribution = product of marginal distributions
  - Marginals are easier to estimate from a small sample

NPTEL

Naïve Bayes Classifier

- Assume independence among attributes A_i when class is given:
  - $P(A_1, A_2, \ldots, A_n \mid C) = P(A_1 \mid C_j)\, P(A_2 \mid C_j) \ldots P(A_n \mid C_j)$
  - Can estimate $P(A_i \mid C_j)$ for all A_i and C_j.
  - New point is classified to C_j if $P(C_j) \prod P(A_i \mid C_j)$ is maximal.

NPTEL

Example of Naïve Bayes Classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

A: attributes

M: mammals

N: non-mammals

$$P(A \mid M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A \mid N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A \mid M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A \mid N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

NPTEL

Naïve Bayes Classifier: Smoothing

- If one of the conditional probability is zero, then the entire expression becomes zero
- Probability estimation:

$$\text{Original} : P(A_i \mid C) = \frac{N_{ic}}{N_c}$$

$$\text{Laplace} : P(A_i \mid C) = \frac{N_{ic} + 1}{N_c + c}$$

$$\text{m - estimate} : P(A_i \mid C) = \frac{N_{ic} + mp}{N_c + m}$$

c: number of classes

p: prior probability

m: parameter

NPTEL

Conditional Independence

- Event A and B are ***conditionally independent given C*** in case

$$\Pr(AB|C)=\Pr(A|C)\Pr(B|C)$$

- A set of events $\{A_i\}$ is conditionally independent given C in case $\Pr(\bigcup_i A_i \mid C) = \prod_i \Pr(A_i \mid C)$

NPTEL

CI: Conditional Independence

- Variables are rarely independent but we can still leverage local structural properties like CI.
- $X \perp Y \mid Z$ if once Z is observed, knowing the value of Y does not change our belief about X
  - The following should hold for all x,y,z
  - $P(X=x \mid Z=z, Y=y) = P(X=x \mid Z=z)$
  - $P(Y=y \mid Z=z, X=x) = P(Y=y \mid Z=z)$
  - $P(X=x, Y=y \mid Z=z) = P(X=x \mid Z=z)\, P(Y=y \mid Z=z)$

NPTEL

Example

Let the two events be the probabilities of persons A and B getting home in time for dinner, and the third event is the fact that a snow storm hit the city. While both A and B have a lower probability of getting home in time for dinner, the lower probabilities will still be independent of each other. That is, the knowledge that A is late does not tell you whether B will be late. (They may be living in different neighborhoods, traveling different distances, and using different modes of transportation.) However, if you have information that they live in the same neighborhood, use the same transportation, and work at the same place, then the two events are NOT conditionally independent.

NPTEL

Exercise: Conditional independence

p(smart ∧ study ∧ prep)	smart		¬smart	
	study	¬study	study	¬study
prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

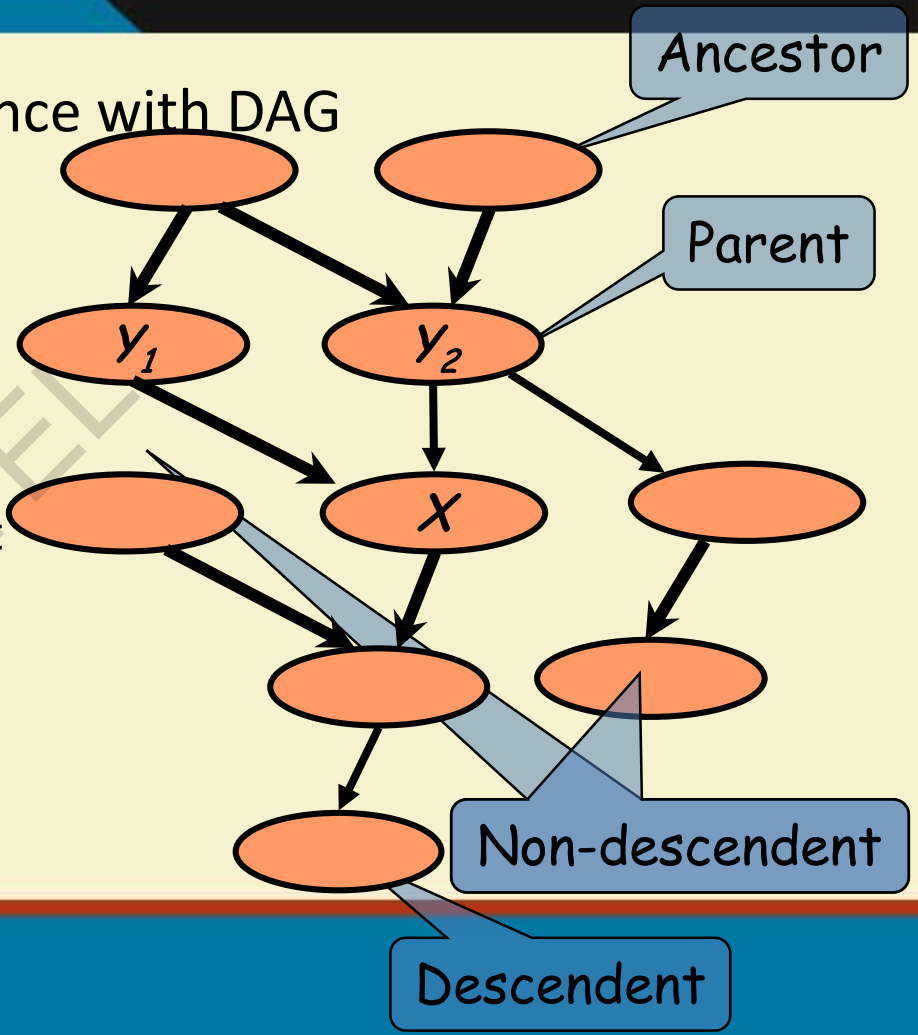
NPTEL

- **Queries:**
  - **Is *smart* conditionally independent of *prepared*, given *study*?**
  - **Is *study* conditionally independent of *prepared*, given *smart*?**

Representing Conditional Independence with DAG

- We now make this independence assumption more precise for **directed acyclic graphs** (DAGs)

- Each random variable X, is independent of its non-descendents, given its parents Pa(X)
- Formally, I (X, NonDesc(X) | Pa(X))



NPTEL

Summary

Advantages:

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes

Drawback:

- Independence assumption may not hold for some attributes
  - Length and weight of a fish are not independent
  - Conditional Independence

NPTEL

End of Bayes Classifier

NPTEL