Machine Learning and Data Mining
Naïve Bayes Classifier

Patricia Hoffman, PhD

Naïve Bayes
Main Sources

- Slides from Tan, Steinbach, Kumar
  - http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap5_alternative_classification.pdf
- Review of Probability
  - http://cs229.stanford.edu/section/cs229-prob.pdf
- R Package with Naïve Bayes
  - R Functions: NaiveBayes() & partimat()
  - http://cran.r-project.org/web/packages/klaR/klaR.pdf
- Project - Bayesian Networks for Data Mining
  - https://online.ucsc-extension.edu/access/content/group/2a4f48bb-81ce-4236-b22b-758d12720b22/Project/Papers/Tutorial-BayesianNetworks.pdf

Naïve Bayes – Example Code

- NaiveBayesClassifier.R
  - Used to Classify the Iris Data
  - NaiveBayes()
  - partimat()

Review of Notation

Logical

$\forall$ for all

$\exists$ there exists

! unique

$\wedge$ and: The statement $A \wedge B$ is true if A and B are both true

$\vee$ or: The statement $A \vee B$ is true if A or B (or both) are true

Set Theory

$\emptyset$ empty or null set

$\cap$ intersection

$\cup$ union

$\subset$ contained in

$\in$ element of

$\notin$ not an element of

$\setminus$ $A \setminus B$ means the set that contains all those elements of A that are not in B.

Axioms of Probabilities

A Probability Measure P[] is a function that maps events in the Sample Space S to real numbers such that

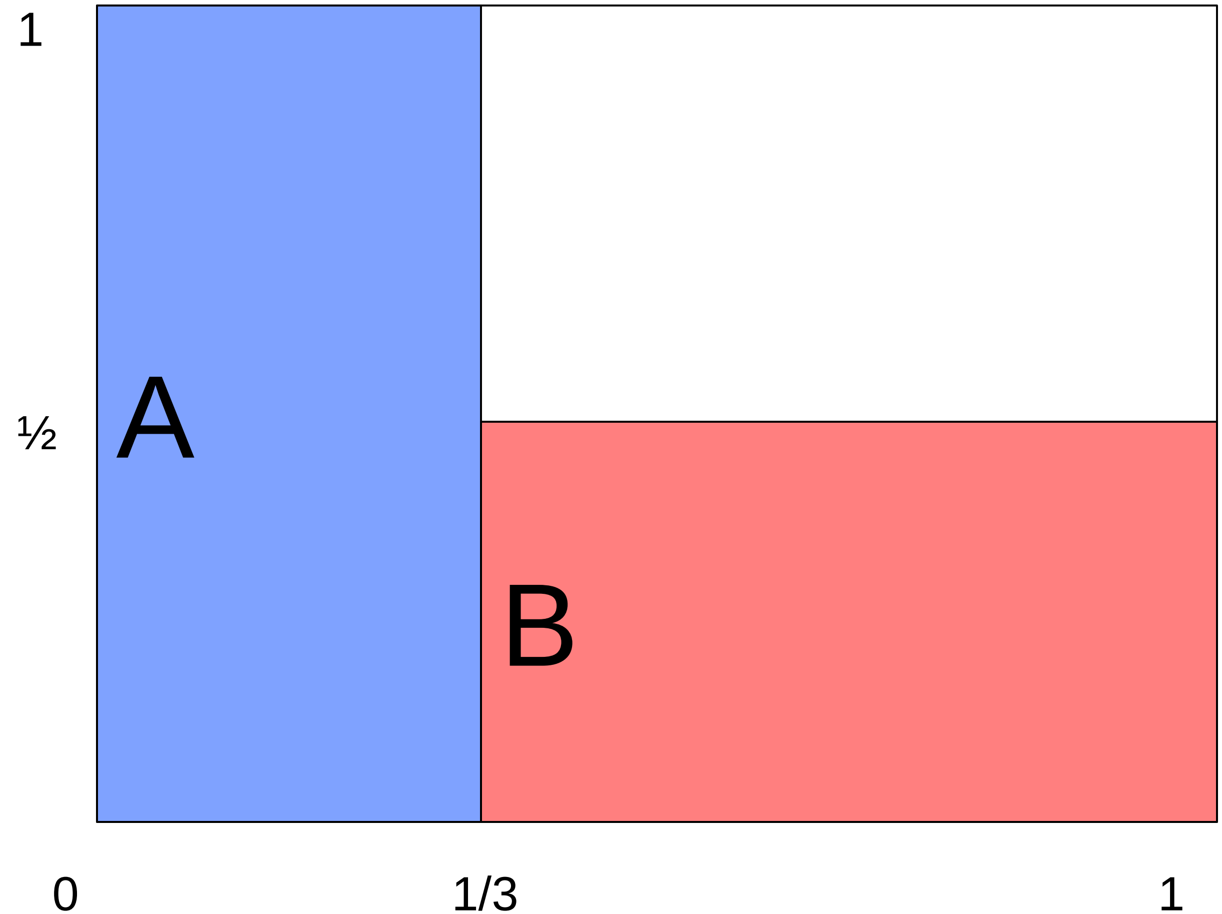
- Axiom 1: For any event A, $P[A] \geq 0$
- Axiom 2: $P[S] = 1$
- Axiom 3: For any countable collection

  A1,A2 , … of mutually exclusive events

  $P[A1 \cup A2 \cup \; ....] = P[A1] + P[A2] + \ldots$

If $A \cap B = \emptyset \Rightarrow P[A \cup B] = P[A] + P[B]$

$P[A] = 1/3$
$P[B] = 1/3$
$P[A \cup B] = 2/3$



Theorems

- $P[\varnothing] = 0$

- $P[A^c] = 1 - P[A]$

  where A^c is the compliment of A

- If $A \subset B \Rightarrow P[A] \leq P[B]$

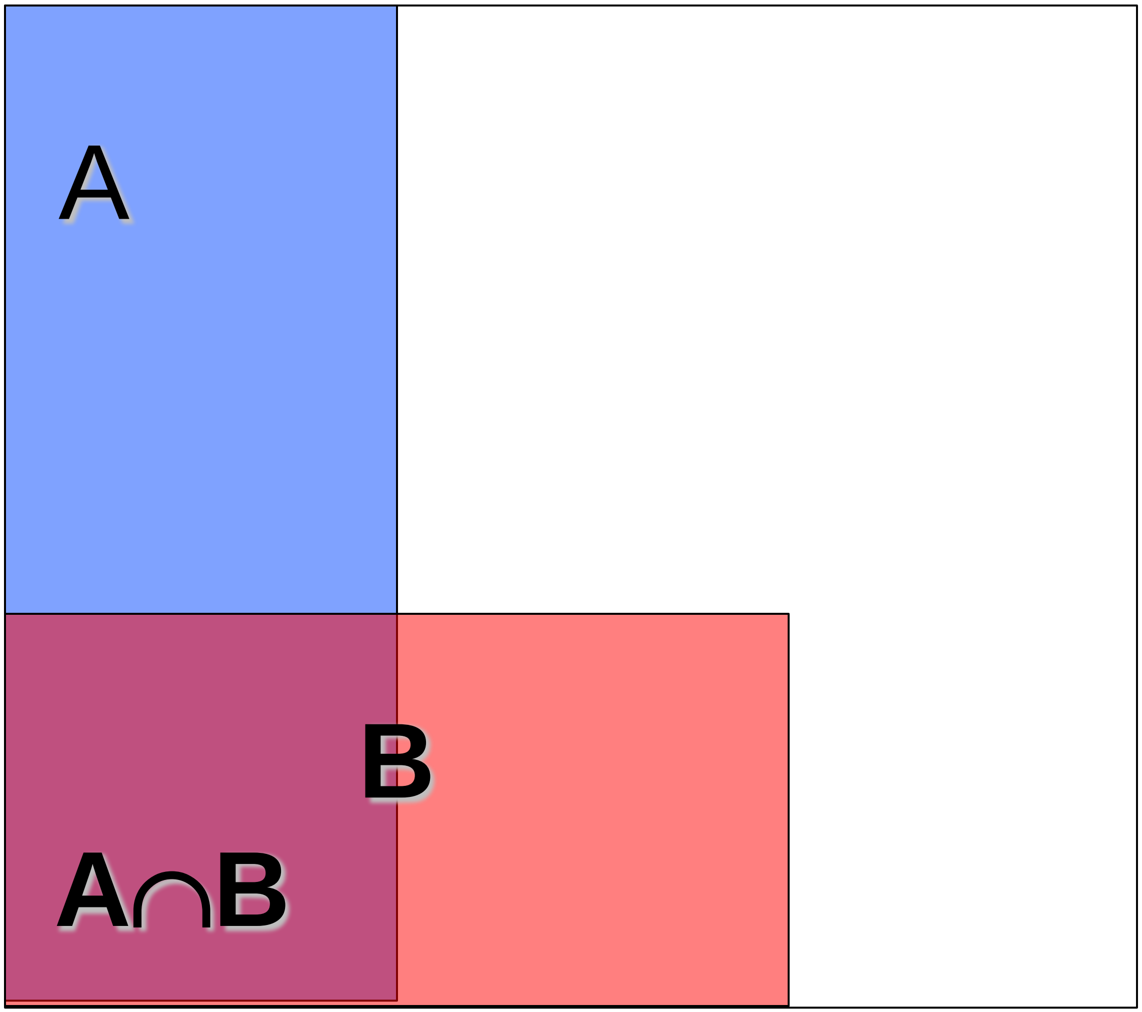
Conditional Probability

$P[A|B]$ = Conditional probability describes our knowledge of A when we know that B has occurred but we still don't know the precise outcome

Conditional Probability

only defined if $P[B] \neq 0$

$P[A|B] =$
$P[A \cap B]/P[B]$



$1/2 =$
$(1/9) / 2/9$

Develop Intuition for Bayes Rule

Given $P[B] \neq 0$

then

$P[A|B] \quad = \quad P[A \cap B]/P[B]$

Multiply both sides by $P[B]$

$P[A \cap B] \quad = \quad P[A|B]P[B]$

But the same is true if $P[A] \neq 0$

then

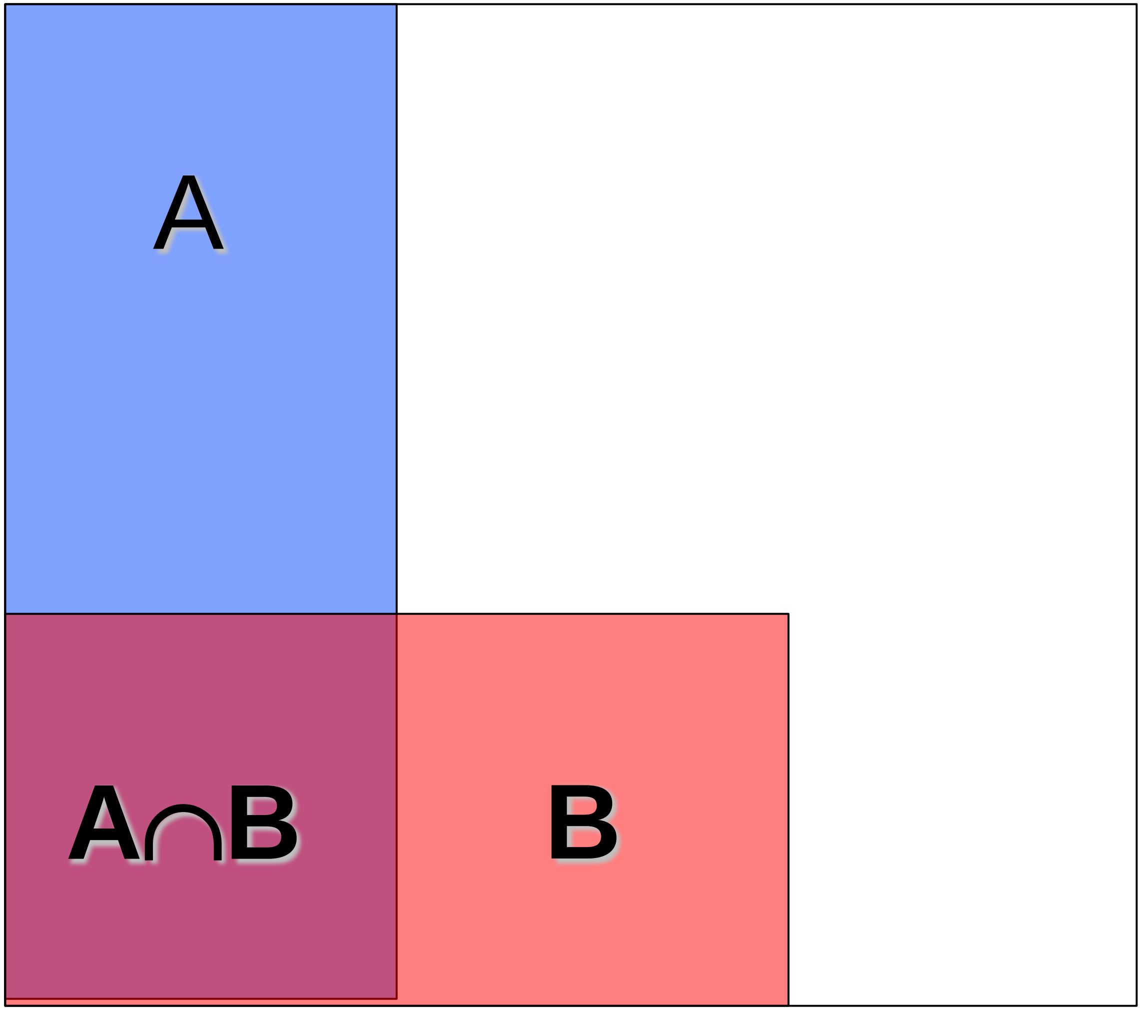
$\mathbf{P[B|A]} \quad = \quad P[A \cap B]/P[A]$

$$\mathbf{= P[A|B]P[B] / P[A]}$$

Bayes Rule (P[B] ≠ 0)

$P[A \cap B] =$
$P[A|B]P[B]$

$1/9 =$
$\frac{1}{2} \times 2/9$



Bayes Rule

Given $P[A] \neq 0$, then

$P[B|A] = P[A|B]P[B] / P[A]$

Example of Bayes Theorem

- Given:
  - A doctor knows that meningitis causes stiff neck 50% of the time
  - Prior probability of any patient having meningitis is 1/50,000
  - Prior probability of any patient having stiff neck is 1/20

- If a patient has stiff neck, what's the probability he/she has meningitis?

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

- Because the patient has a stiff neck, the probability that the patient has meningitis has gone up by a factor of 10

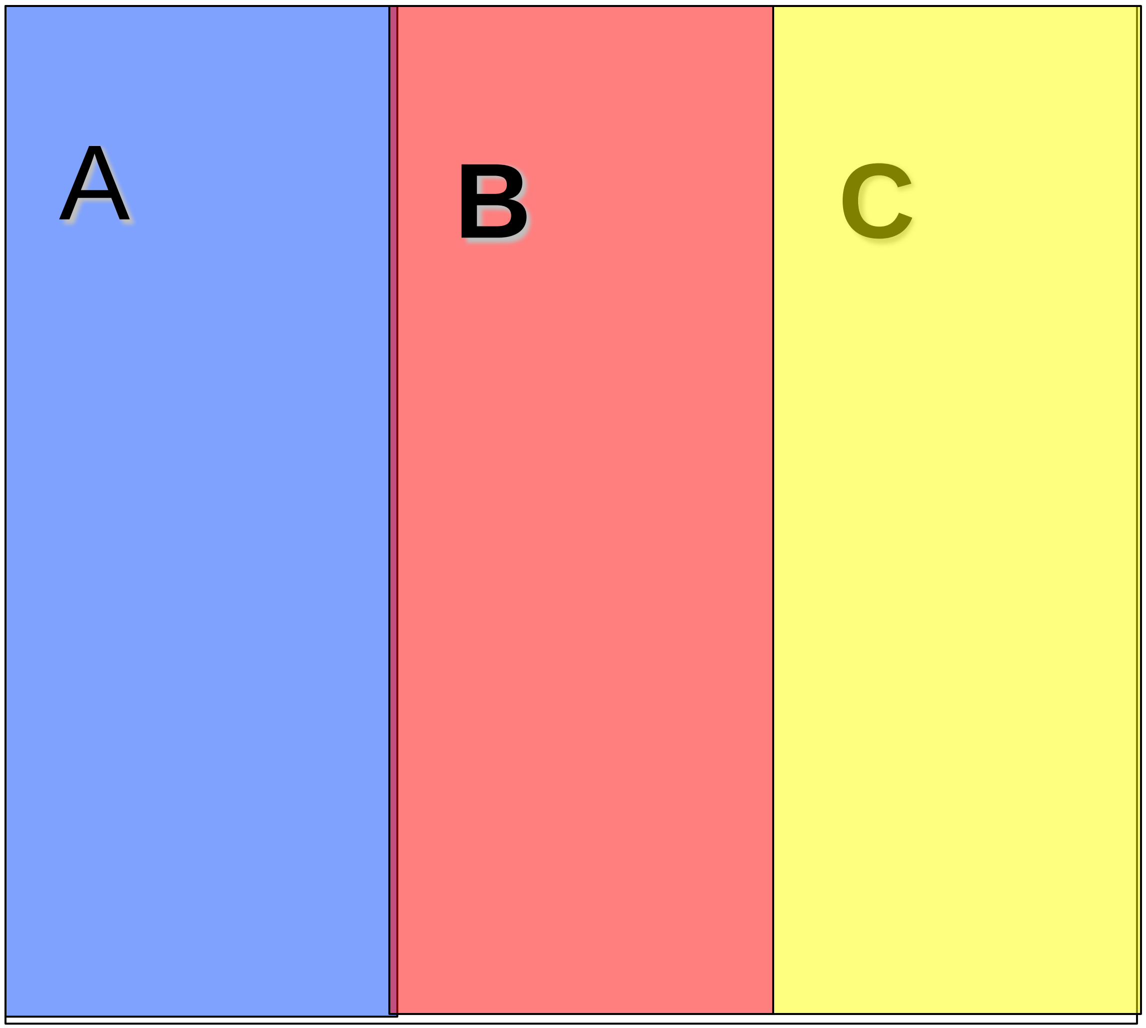
Lets Make a Deal

Three Doors A, B, C

- One Grand Prize … Car

Contestant picks a door. Monet Hall opens one of the other doors … it is a booby prize

Should Contestant stay with original door or switch?


A
B
C

Contestant picks door A

Monty opens door B;

Should contestant switch to door C?

Solution: Compare Probabilities

grand prize behind door A with

grand prize behind door C

Preliminaries:

Probability(Monty opens door B) = P(oB)

Probability(Car behind door X) = P(X)

Answer: Switch …
Probability Car behind door C is greater

$$\mathrm{P}(\text{Monty opens door B}) = \mathrm{P}(\mathrm{oB})$$

$$\begin{aligned}
&= \mathrm{P}(\mathrm{oB}|\mathrm{A})\mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{oB}|\mathrm{B})\mathrm{P}(\mathrm{B}) + \mathrm{P}(\mathrm{oB}|\mathrm{C})\mathrm{P}(\mathrm{C}) \\
&= (1/2)(1/3) + (0)(1/3) + (1)(1/3) \\
&= 1/2
\end{aligned}$$

$$\mathrm{P}(\text{car in A}|\text{open B}) = \frac{\mathrm{P}(\mathrm{oB}|\mathrm{A})\mathrm{P}(\mathrm{A})}{\mathrm{P}(\mathrm{oB})} = \frac{(1/2)(1/3)}{1/2} = 1/3$$

$$\mathrm{P}(\text{car in C}|\text{open B}) = \frac{\mathrm{P}(\mathrm{oB}|\mathrm{C})\mathrm{P}(\mathrm{C})}{\mathrm{P}(\mathrm{oB})} = \frac{(1)(1/3)}{1/2} = 2/3$$

Intuitive Solution

Think of it as two choices

Choice 1) Door A has a probability of 1/3 having the car behind it. (as each of the doors are equally likely)

Choice 2) Not Door A has a probability of 2/3. Monty puts the full 2/3 probability on the exact door to switch over to.

Bayesian Classifiers

- Consider each attribute and class label as random variables
- Given a record with attributes (A_1, A_2,…,A_n)
  - Goal is to predict class C
  - Specifically, we want to find the value of C that maximizes $P(C| A_1, A_2,\ldots,A_n)$
- Can we estimate $P(C| A_1, A_2,\ldots,A_n)$ directly from data?

Bayesian Classifiers

- Approach:
  - compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of C using the Bayes theorem

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

  - Choose value of C that maximizes
    $P(C \mid A_1, A_2, \ldots, A_n)$
  - Equivalent to choosing value of C that maximizes
    $P(A_1, A_2, \ldots, A_n \mid C)\, P(C)$

- How to estimate $P(A_1, A_2, \ldots, A_n \mid C)$?

Conditional Independence

Consider three Random Variables: A, B, and C

A is conditionally independent of B, given C, if

$$P[A|B \cap C] = P[A|C]$$

Example:

Reading ability may depend on arm length, however if age is fixed then reading ability is independent of reading ability

Naïve Bayes Classifier

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

How to Estimate Probabilities from Data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Class: $P(C) = N_c / N$
  - e.g., P(No) = 7/10, P(Yes) = 3/10
- For discrete attributes:

$$P(A_i \mid C_k) = |A_{ik}| / N_c$$

  - where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k
  - Examples:

P(Status=Married|No) = 4/7
P(Refund=Yes|Yes)=0

How to Estimate Probabilities from Data?

- For continuous attributes:
  - Discretize the range into bins
    - one ordinal attribute per bin
    - violates independence assumption
  - Two-way split: (A < v) or (A > v)
    - choose only one of the two splits as new attribute
  - Probability density estimation:
    - Assume attribute follows a normal distribution
    - Use data to estimate parameters of distribution (e.g., mean and standard deviation)
    - Once probability distribution is known, can use it to estimate the conditional probability $P(A_i|C)$

How to Estimate Probabilities from Data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Normal distribution:

$$P(A_i \mid c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

  – One for each (A_i, c_i) pair

- For (Income, Class=No):
  – If Class=No
    - sample mean = 110
    - sample variance = 2975

$$P(Income = 120 \mid No) = \int_{120}^{120+\varepsilon} \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(x-110)^2}{2(2975)}} dx \cong 0.0072\varepsilon$$

The constant ε appears in each and as we are trying to find the maximum only 0.0072 is considered

Example of Naïve Bayes Classifier

Given a Test Record:

$$X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$$

naive Bayes Classifier:

P(Refund=Yes|No) = 3/7
P(Refund=No|No) = 4/7
P(Refund=Yes|Yes) = 0
P(Refund=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No)=1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/7
P(Marital Status=Divorced|Yes)=1/7
P(Marital Status=Married|Yes) = 0

For taxable income:

If class=No:	sample mean=110	
	sample variance=2975	
If class=Yes:	sample mean=90	
	sample variance=25	

- P(X|Class=No) = P(Refund=No|Class=No) $\times$ P(Married| Class=No) $\times$ P(Income=120K| Class=No) = 4/7 $\times$ 4/7 $\times$ 0.0072 = 0.0024

- P(X|Class=Yes) = P(Refund=No| Class=Yes) $\times$ P(Married| Class=Yes) $\times$ P(Income=120K| Class=Yes) = 1 $\times$ 0 $\times$ 1.2 $\times$ 10^{-9} = 0

Since P(X|No)P(No) > P(X|Yes)P(Yes)

Therefore P(No|X) > P(Yes|X)

=> Class = No

Naïve Bayes Classifier

- 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

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

A: attributes

M: mammals

N: non-mammals

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

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

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

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

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

Naïve Bayes (Summary)

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes
- Independence assumption may not hold for some attributes
  – Use other techniques such as Bayesian Belief Networks (BBN)

Iris Data: 50 samples from each of three species Setosa, Versicolor, Virginica

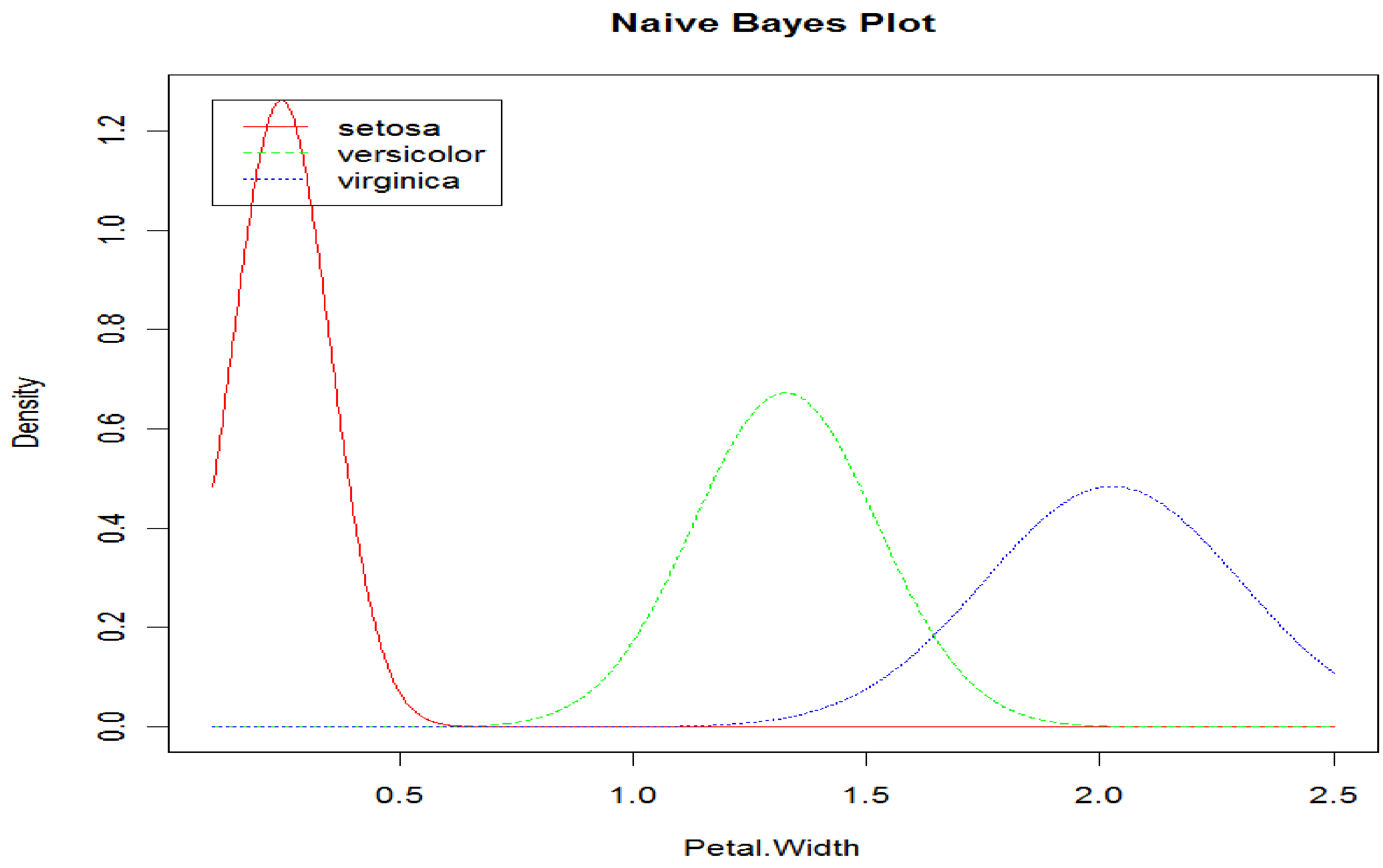
5 columns of data:
sepal length, sepal width, petal length, sepal width, species

[Iris Species - Versicolor]

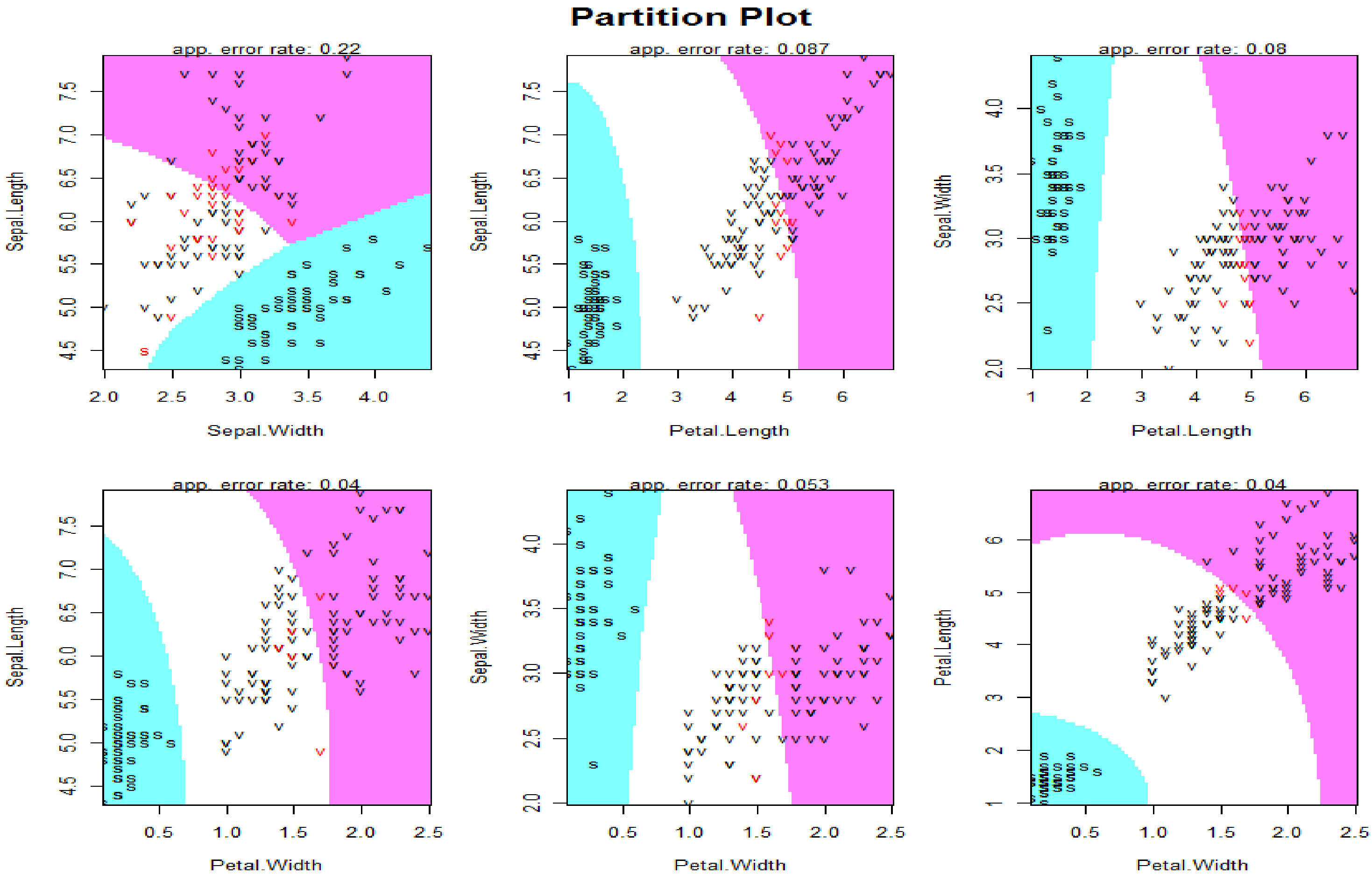
Naïve Bayes – Example Code

- NaiveBayesClassifier.R
  – Used to Classify the Iris Data
  – NaiveBayes()
  – partimat()

Distributions for Petal Width



Partition Plot



Error Comparisons

Ridge Regression Multi-Class

(linear basis function)

Error 43/150 = 0.287

Much Better

NaiveBayes(Species ~ ., data = iris)

Error 0.04