Chapter 5

Naïve Bayes

Naïve Bayes

- ***Naïve Bayes*** is best suited for ML applications…
  - wherein the feature vectors $\bar{\mathbf{x}}$ are DISCRETE.

Naïve Bayes

- ***Naïve Bayes*** is best suited for ML applications… Response is Discrete
  - wherein the feature vectors $\bar{\mathbf{x}}$ are **discrete**

Naïve Bayes

- Example: ML-based e-mail **spam filter**

Genuine

Spam

Naïve Bayes

- Consider a **feature vector** $\bar{\mathbf{x}}$ of size N
  - where N is the number of words in the

English Language Dictionary

$N \approx 60{,}000$

Feature vector

Large dimensional Vector

- For each e-mail we create a **feature vector** $\bar{\mathbf{x}}$ of size N
  - where N is the number of words in the **English language dictionary**

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_{60,000} \end{bmatrix} = 60{,}000 \times 1$$

Naïve Bayes

- The entry $x_j = 1$, if the ***email*** contains the j −th word of the ***dictionary***,…
  - else $x_j = 0$

1000^{th} word in dictionary = "able"

$x_{1000} = 1$ if "able" occurs in e-mail

else $x_{1000} = 0$

Naïve Bayes

$$\bar{\mathbf{x}} = \begin{bmatrix} \vdots \\ x_{1000} \\ x_{1001} \\ x_{1002} \\ x_{1003} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ 1 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix} \begin{array}{l} \vdots \\ \text{able} \\ \text{above} \\ \text{abroad} \\ \text{access} \\ \vdots \end{array}$$

occurs.

Does NOT occur

Does NOT occur

occurs

Feature Vector

I am able to access . . . . .

Naïve Bayes

- The labels $y = 1{,}0$ indicate SPAM, GENUINE. e-mails, respectively

Naïve Bayes

- The labels $y = 1,0$ indicate **spam, genuine** e-mails, respectively

SPAM $\Rightarrow y = 0$

GENUINE $\Rightarrow y = 1.$

Naïve Bayes

- **Naïve Bayes assumption**:
- The different words are Conditionally independent given the label y

Naïve Bayes

- **Naïve Bayes assumption**:
- The different words are **conditionally independent** given the label y

Naïve Bayes

$v_i \in \{0, 1\}$

$u \in \{0, 1\}$

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = u)$$

$$= p(x_1 = v_1, x_2 = v_2, \ldots, x_N = v_N | y = u)$$

$$= \underline{p(x_1 = v_1 | y = u) \times p(x_2 = u_2 | y = u) \times \cdots \times}$$

INDEPENDENCE

$$\underline{\cdots \times p(x_N = v_N | y = u)}$$

$$= \prod_{i=1}^{N} p(x_i = v_i | y = u)$$

Naïve Bayes

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = u)$$

Joint Probability

$$= p(x_1 = v_1, \dots, x_N = v_N | y = u)$$

$$= p(x_1 = v_1 | y = u) \times \cdots \times p(x_N = v_N | y = u)$$

marginal Probabilities

$$= \prod_{j=1}^{N} p(x_j = v_j | y = u)$$

independence

Naive Bayes.

$P(A \cap B) = P(A) \cdot P(B)$

if A, B are independent

Naïve Bayes

- The quantities $p(x_j = v_j | y = u)$ are the Prior Probabilities.
  - How to calculate these?

Probabilities of word occurences in spam/Genuine e-mails.

Naïve Bayes

- The quantities $p\left(x_j = v_j | y = u\right)$ are the **prior probabilities**
  - How to calculate these?

Naïve Bayes

- Consider the availability of M ***training pairs*** $\left(\bar{\mathbf{x}}(i), y(i)\right)$

labels

Feature vectors

$\bar{x}(1), y(1)$

$\bar{x}(2), y(2)$

$\vdots$

$\bar{x}(M), y(M)$

Training set

Naïve Bayes

- The various ***prior probabilities*** can now be calculated as follows

Naïve Bayes

Number of spam e-mails with j^{th} word.

$p(x_j = 1 | y = 1) =$ Probability j^{th} word occurs in spam e-mail

$$= \frac{\#\text{ spam emails with } j^{th} \text{ word}}{\text{Total } \#\text{ spam e-mails}}.$$

$$= \frac{\sum_{i=1}^{M} 1(x_j(i) = 1, y(i) = 1)}{\sum_{i=1}^{M} 1(y(i) = 1)}$$

indicator function

= Total # spam e-mails.

Naïve Bayes

$$p(x_j = 1|y = 1)$$

$$= \frac{\text{Number of spam e} - \text{mails with jth word}}{\text{Total number of spam e} - \text{mails}}$$

$$= \frac{\sum_{i=1}^{M} \mathbf{1}(x_j(i) = 1, y(i) = 1)}{\sum_{i=1}^{M} 1(y(i) = 1)}$$

conditional prior probability
jth word occurs in a spam e-mail.

Naïve Bayes

$$p(x_j = 1 | y = 0) = \text{Probability } j^{th} \text{ word occurs in genuine e-mail}$$

$$= \frac{\text{number of genuine e-mails with } j^{th} \text{ word}}{\text{Total number of genuine e-mails.}}$$

$$= \frac{\sum_{i=1}^{M} 1\left(x_j(i) = 1,\ y(i) = 0\right)}{\sum_{i=1}^{M} 1\left(y(i) = 0\right)}$$

Naïve Bayes

Training data.

Prior probabity.

$$p\left(x_j = 1 | y = 0\right) = \frac{\sum_{i=1}^{M} \mathbf{1}\left(x_j(i) = 1, y(i) = 0\right)}{\sum_{i=1}^{M} 1\left(y(i) = 0\right)}$$

Naïve Bayes

$p(y = 1) =$ Probability of Spam email.

$$= \frac{\text{Number of spam e-mails}}{\text{Total number of e-mails}}$$

$$= \frac{\sum_{i=1}^{M} 1(y(i) = 1)}{M}$$

Naïve Bayes

$$p(y=1) = \frac{\sum_{i=1}^{M} \mathbf{1}(y(i)=1)}{M}$$

Prior probability of spam e-mail.

Naïve Bayes

Prob. j^{th} word is absent in spam e-mail

$$p(x_j = 0 | y = 1) = 1 - p(x_j = 1 | y = 1).$$

Prob jth word is absent in genuine e-mail

$$p(x_j = 0 | y = 0) = 1 - p(x_j = 1 | y = 0).$$

Naïve Bayes

$$p(x_j = 0 | y = 1) = 1 - p(x_j = 1 | y = 1)$$

$$p(x_j = 0 | y = 0) = 1 - p(x_j = 1 | y = 0)$$

Naïve Bayes

$$p(y = 0) = 1 - p(y=1)$$

Probability of genuine e-mail

Probability of spam e-mail.

Naïve Bayes

$$p(y = 0) = 1 - p(y = 1)$$

Naïve Bayes

$\bar{x} = \bar{V}$: Feature vector of e-mail.

- Finally, the **posterior probabilities** are calculated

Posterior Probabilities.

$$p(y = 1|\bar{\mathbf{x}} = \bar{\mathbf{v}})$$

Probability of spam e-mail given $\bar{x}$

$$p(y = 0|\bar{\mathbf{x}} = \bar{\mathbf{v}})$$

Probability of genuine e-mail given $\bar{x}$

Naïve Bayes

Bayes rule

- We use the principle

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

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

$$\Rightarrow p(B|A) = \frac{p(A|B)\,p(B)}{p(A)}.$$

Naïve Bayes

- We use the principle

$$\underbrace{p(B|A) = \frac{p(A|B)p(B)}{p(A)}}_{\text{Bayes Rule.}}$$

Naïve Bayes

Prior probabilities

$$p(\overset{B}{y} = 1|\overset{A}{\bar{\mathbf{x}}} = \bar{\mathbf{v}}) = \frac{p(\bar{x}=\bar{v}|y=1)\,p(y=1)}{p(\bar{x}=\bar{v})}$$

Prior probabilities.

$$p(y = 0|\bar{\mathbf{x}} = \bar{\mathbf{v}}) = \frac{p(\bar{x}=\bar{v}|y=0)\,p(y=0)}{P(\bar{x}=\bar{v})}$$

Naïve Bayes

$$p(y = 1|\bar{\mathbf{x}} = \bar{\mathbf{v}}) = \frac{p(\bar{\mathbf{x}} = \bar{\mathbf{v}}|y = 1) \times p(y = 1)}{p(\bar{\mathbf{x}} = \bar{\mathbf{v}})}$$

e-mail is Spam

$$p(y = 0|\bar{\mathbf{x}} = \bar{\mathbf{v}}) = \frac{p(\bar{\mathbf{x}} = \bar{\mathbf{v}}|y = 0) \times p(y = 0)}{p(\bar{\mathbf{x}} = \bar{\mathbf{v}})}$$

e-mail is genuine

Aposteriori Probabilities.

Naïve Bayes

- E-mail is classified as ***spam*** if

$$p(y=1 \mid \bar{x}=\bar{v}) > p(y=0 \mid \bar{x}=\bar{v})$$

$\Longrightarrow$ classified as spam

$$p(y=1 \mid \bar{x}=\bar{v}) \leq p(y=0 \mid \bar{x}=\bar{v}) \Longrightarrow \text{classified as genuine.}$$

Naïve Bayes

- E-mail is classified as ***spam*** if
  $p(y=1|\bar{\mathbf{x}}=\bar{\mathbf{v}}) > p(y=0|\bar{\mathbf{x}}=\bar{\mathbf{v}})$

Naïve Bayes

$$p(y = 1|\bar{\mathbf{x}} = \bar{\mathbf{v}}) > p(y = 0|\bar{\mathbf{x}} = \bar{\mathbf{v}})$$

$$\Rightarrow \frac{p(\bar{x}=\bar{v}|y=1)\,p(y=1)}{\cancel{p(\bar{x}=\bar{v})}} > \frac{p(\bar{x}=\bar{v}|y=0)\,p(y=0)}{\cancel{p(\bar{x}=\bar{v})}}$$

$$\Rightarrow \underbrace{p(\bar{x}=\bar{v}|y=1)\,p(y=1)}_{\alpha_1} > \underbrace{p(\bar{x}=\bar{v}|y=0)\,p(y=0)}_{\alpha_0}.$$

Naïve Bayes

$$p(y=1|\bar{\mathbf{x}}=\bar{\mathbf{v}}) > p(y=0|\bar{\mathbf{x}}=\bar{\mathbf{v}})$$

$$\Rightarrow \frac{p(\bar{\mathbf{x}}=\bar{\mathbf{v}}|y=1)\times p(y=1)}{p(\bar{\mathbf{x}}=\bar{\mathbf{v}})} > \frac{p(\bar{\mathbf{x}}=\bar{\mathbf{v}}|y=0)\times p(y=0)}{p(\bar{\mathbf{x}}=\bar{\mathbf{v}})}$$

$$\Rightarrow \underbrace{\boxed{p(\bar{\mathbf{x}}=\bar{\mathbf{v}}|y=1)\times p(y=1)}}_{Q_1}$$

$$> \underbrace{\boxed{p(\bar{\mathbf{x}}=\bar{\mathbf{v}}|y=0)\times p(y=0)}}_{Q_0}.$$

$Q_1 > Q_0 \Rightarrow$ Spam

$Q_1 < Q_0 \Rightarrow$ Genuine

Naïve Bayes

- Note

Naive Bayes: ~~independence~~ conditional independence of words.

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = 1) = \prod_{j=1}^{N} p(x_j = v_j | y = 1)$$

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = 0) = \prod_{j=1}^{N} p(x_j = v_j | y = 0).$$

Naïve Bayes

- Note

Naive Assumption

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = 1) = \prod_{j=1}^{N} p\left(x_j = v_j | y = 1\right)$$

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = 0) = \prod_{j=1}^{N} p\left(x_j = v_j | y = 0\right)$$

Naïve Bayes

- This implies, E-mail classified as spam if. $\alpha_1 > \alpha_0$ else Genuine.

$$p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = 1) \times p(y = 1) > p(\bar{\mathbf{x}} = \bar{\mathbf{v}} | y = 0) \times p(y = 0)$$

$$\Rightarrow \underbrace{\prod_{j=1}^{N} p(x_j = v_j | y = 1)\, p(y=1)}_{\alpha_1}$$

$$> \underbrace{\prod_{j=1}^{N} p(x_j = v_j | y = 0)\, p(y=0)}_{\alpha_0}$$

Naïve Bayes

- This implies, E-mail classified as spam if

Likelihoods.

$$\underbrace{\prod_{j=1}^{N} p(x_j = v_j | y = 1) \times p(y = 1)}_{Q_1} > \underbrace{\prod_{j=1}^{N} p(x_j = v_j | y = 0) \times p(y = 0)}_{Q_0}$$

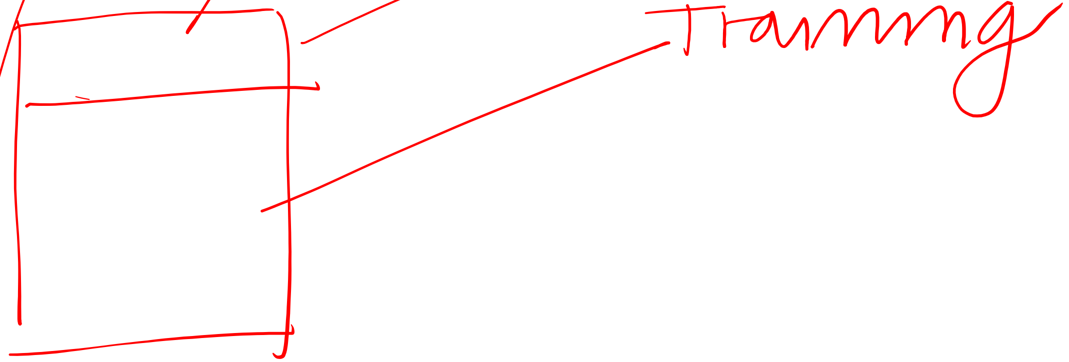
Prior Probabilities.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
```

ScikitLearn

To split data into test & Training

Pandas: Data Handling



Naive Bayes

```
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score
```

For plotting

Scale Data
To bring onto same scale

Accuracy score
Evaluate Performance

CSV File

```
purchaseData = pd.read_csv('Purchase_Logistic.csv')
```

- This dataset contains
- User ID
- Gender

M/F

- Age

Age of customer

- Estimated Salary

Salary

- Purchased column - Has data as 0 and 1

Not Purchase

Purchase

- where 1 denotes that item is purchased.

Purchase_Logistic - Excel

A1 | User ID

User ID	Gender	Age	EstimatedSalary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	0
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	0
15728773	Male	27	58000	0
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	0
15727311	Female	35	65000	0
15570769	Female	26	80000	0
15606274	Female	26	52000	0
15746139	Male	20	86000	0
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1
15617482	Male	45	26000	1
15704583	Male	46	28000	1
15621083	Female	48	29000	1
15649487	Male	45	22000	1
15736760	Female	47	49000	1
15714658	Male	48	41000	1
15599081	Female	45	22000	1
15705113	Male	46	23000	1
15631159	Male	47	20000	1
15792818	Male	49	28000	1
15633531	Female	47	30000	1
15744529	Male	29	43000	0

Purchase_Logistic

User ID :
Gender
Age
Salary
Purchase / NOT Purchase

```
X = purchaseData.iloc[:, [2, 3]].values
Y = purchaseData.iloc[:, 4].values
```

Binary Response

$\in \{0, 1\}$.

Age

Estimated Salary

```
scaler = StandardScaler();
X = scaler.fit_transform(X)

Xtrain, Xtest, Ytrain, Ytest \
= train_test_split(X, Y, test_size = 0.20, random_state = 7)
```

20% of Data for test
80% For training

Removes mean
scales to unit variance

Scale data
To bring salary, age
on same scale

```
cf = GaussianNB()
cf.fit(Xtrain, Ytrain)
Ypred = cf.predict(Xtest)
```

Fit Naive Bayes To Training data.

Naives Bayes Algorithm

Predict Response For Test data.

```
NBscore = accuracy_score(Ypred,Ytest)
print('Accuracy score of NB Classifier is',100*NBscore,'%\n')
```

Prediction

Original
True Response

% correct Response

```
Accuracy score of NB Classifier is 93.75 %

In [20]:
```

≈ 94 %

```python
plt.figure(1);
plt.scatter(X[:, 0], X[:, 1], c = Y)
plt.suptitle('Purchase Data')
plt.xlabel('Scaled Age')
plt.ylabel('Scaled Income')
plt.grid(1,which='both')
plt.axis('tight')
plt.show()
```

Grid

Tight axis.

Limits axis to data.

Age

Salary

Scaled Salary

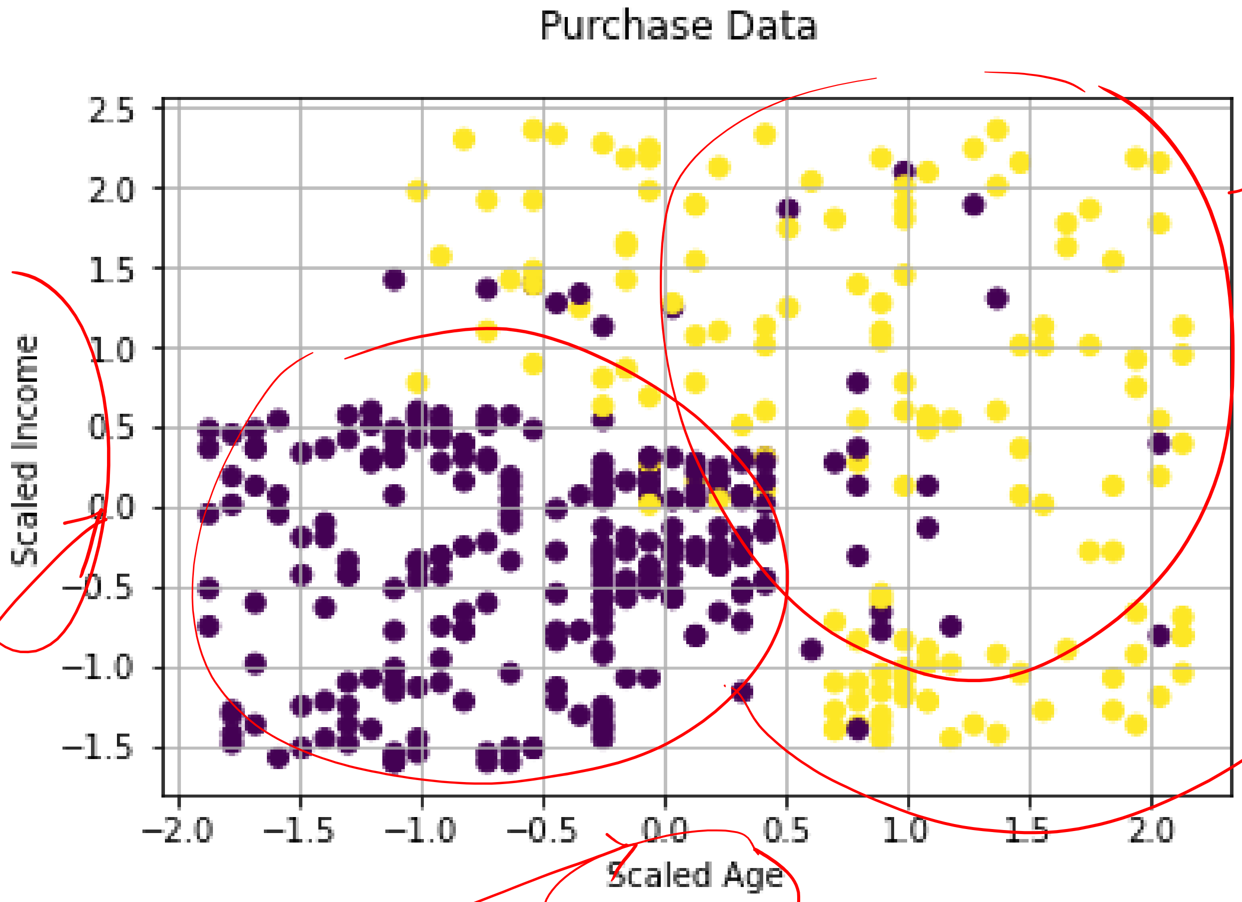
Title.

Xlabel

Color: True Response.

Purchased.

Income.



Not Purchased

Age

centered around 0.

```python
col = cf.predict(X)

plt.figure(2);
plt.scatter(X[:, 0], X[:, 1], c = col)
plt.suptitle('Naive Bayes Purchase Data')
plt.xlabel('Scaled Age')
plt.ylabel('Scaled Income')
plt.grid(1,which='both')
plt.axis('tight')
plt.show()
```

Scaled Age on X axis.

Scaled Salary on y axis.

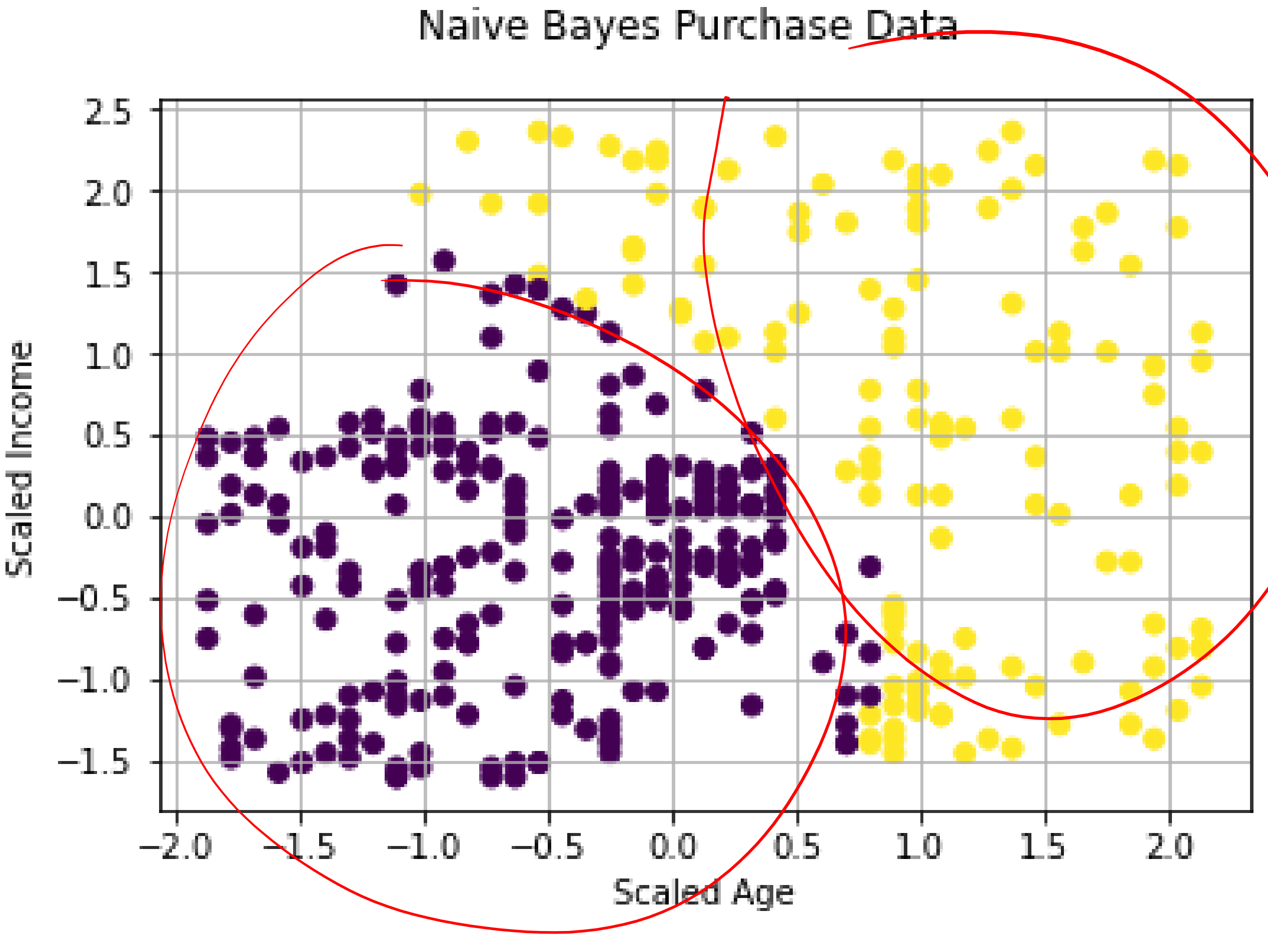
Predicted Response.

Title.

xlabel.

ylabel.

color = Predicted Response.



Predicted Response using Naive Bayes Technique.

Accuracy $\approx 94\%$.

Instructors may use this white area (14.5 cm / 25.4 cm) for the text.
Three options provided below for the font size.

Font: Avenir (Book), Size: 32, Colour: Dark Grey

Font: Avenir (Book), Size: 28, Colour: Dark Grey

Font: Avenir (Book), Size: 24, Colour: Dark Grey

Do not use the space below.