pertemuan 12 pds

2025-04-30

```r
library(naivebayes)
```

```
## naivebayes 1.0.0 loaded
## For more information please visit:
## https://majkamichal.github.io/naivebayes/
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
data(iris)
head(iris)
```

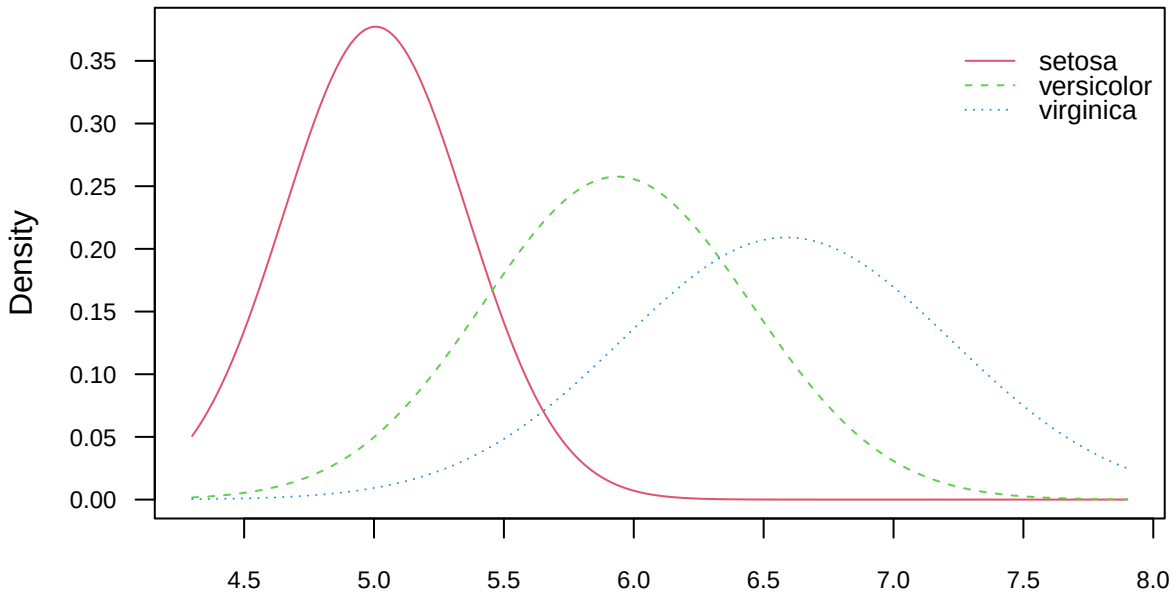
```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

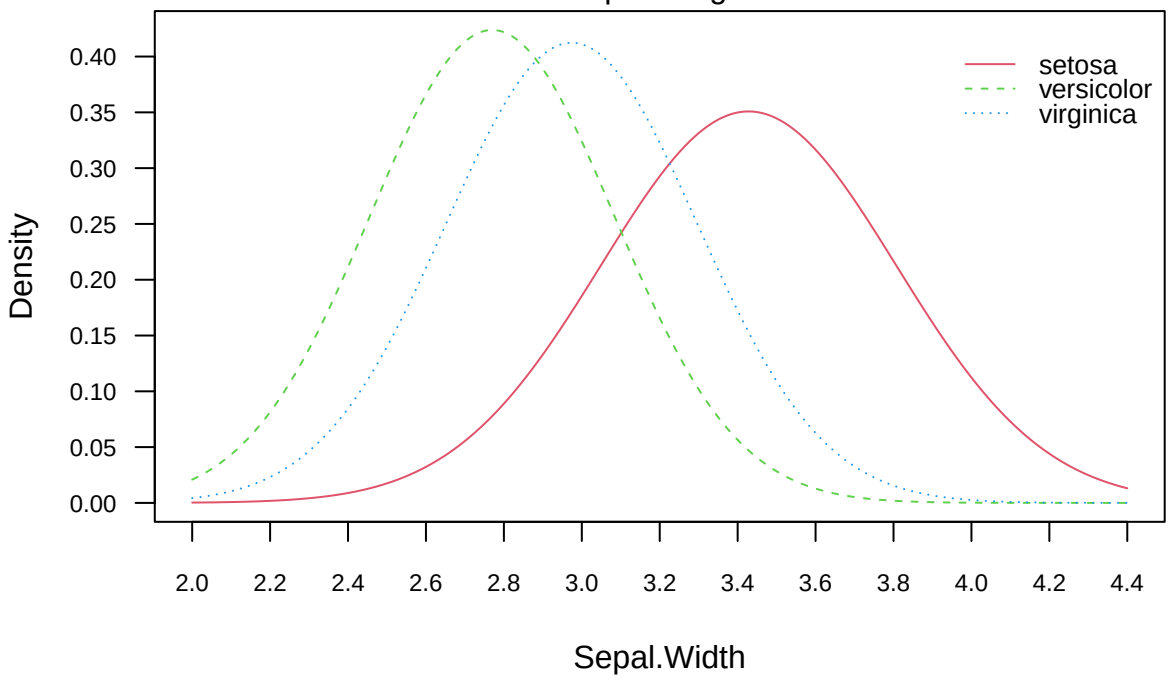
```r
target = "species"
predictor = setdiff(names(iris), target)

model_nb <- naive_bayes(Species ~ ., data = iris)
print(model_nb)
```

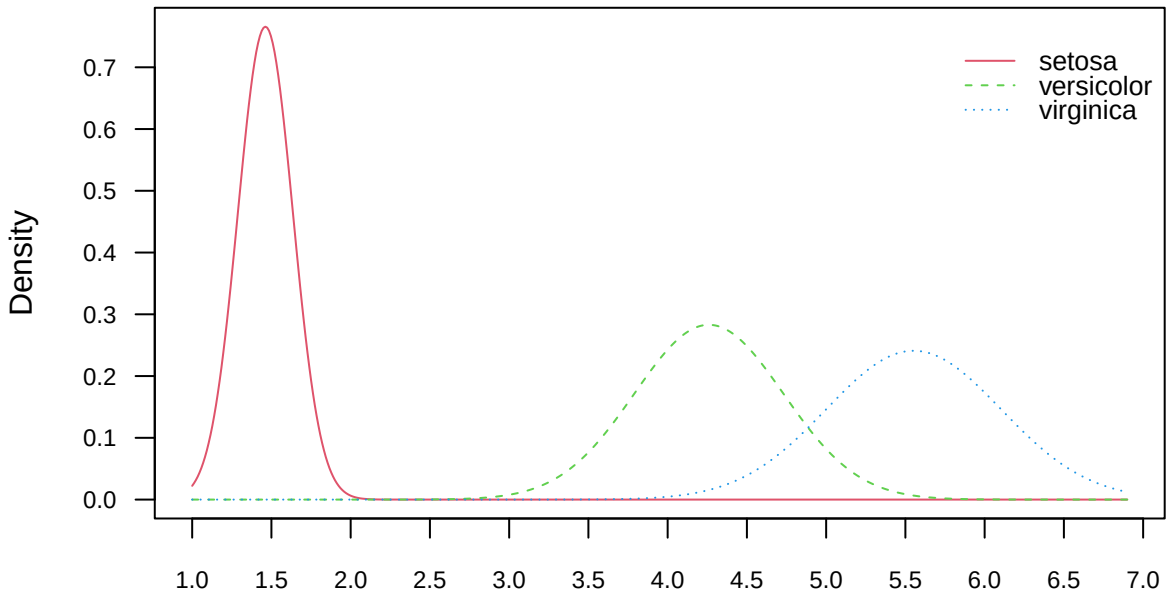
```
##
## ================================== Naive Bayes ==================================
##
## Call:
## naive_bayes.formula(formula = Species ~ ., data = iris)
##
## ---------------------------------------------------------------------------------
##
## Laplace smoothing: 0
##
## ---------------------------------------------------------------------------------
```

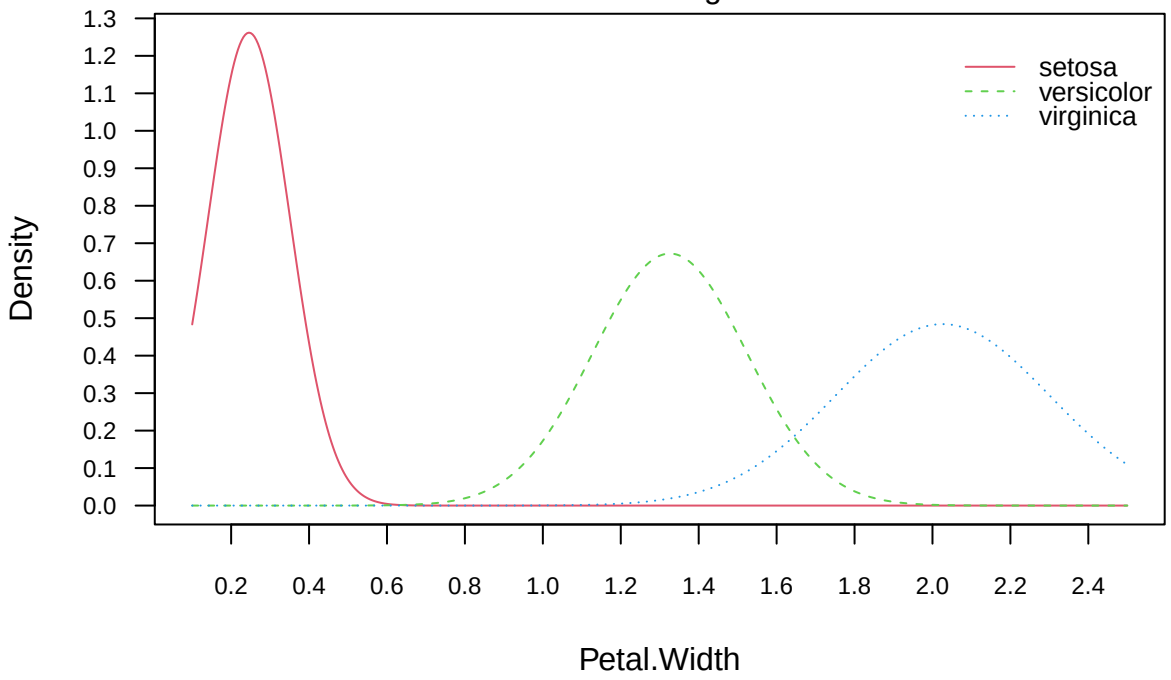
```
##
## A priori probabilities:
##
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
##
## ---------------------------------------------------------------------------
##
## Tables:
##
## ---------------------------------------------------------------------------
## :: Sepal.Length (Gaussian)
## ---------------------------------------------------------------------------
##
## Sepal.Length    setosa versicolor virginica
##         mean 5.0060000  5.9360000 6.5880000
##         sd   0.3524897  0.5161711 0.6358796
##
## ---------------------------------------------------------------------------
## :: Sepal.Width (Gaussian)
## ---------------------------------------------------------------------------
##
## Sepal.Width    setosa versicolor virginica
##        mean 3.4280000  2.7700000 2.9740000
##        sd   0.3790644  0.3137983 0.3224966
##
## ---------------------------------------------------------------------------
## :: Petal.Length (Gaussian)
## ---------------------------------------------------------------------------
##
## Petal.Length    setosa versicolor virginica
##         mean 1.4620000  4.2600000 5.5520000
##         sd   0.1736640  0.4699110 0.5518947
##
## ---------------------------------------------------------------------------
## :: Petal.Width (Gaussian)
## ---------------------------------------------------------------------------
##
## Petal.Width    setosa versicolor virginica
##        mean 0.2460000  1.3260000 2.0260000
##        sd   0.1053856  0.1977527 0.2746501
##
## ---------------------------------------------------------------------------
```

```
plot(model_nb)
```


Density
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
Sepal.Length
setosa
versicolor
virginica


Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2.0
2.2
2.4
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
4.2
4.4
Sepal.Width
setosa
versicolor
virginica






```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(dplyr)
data(mtcars)
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

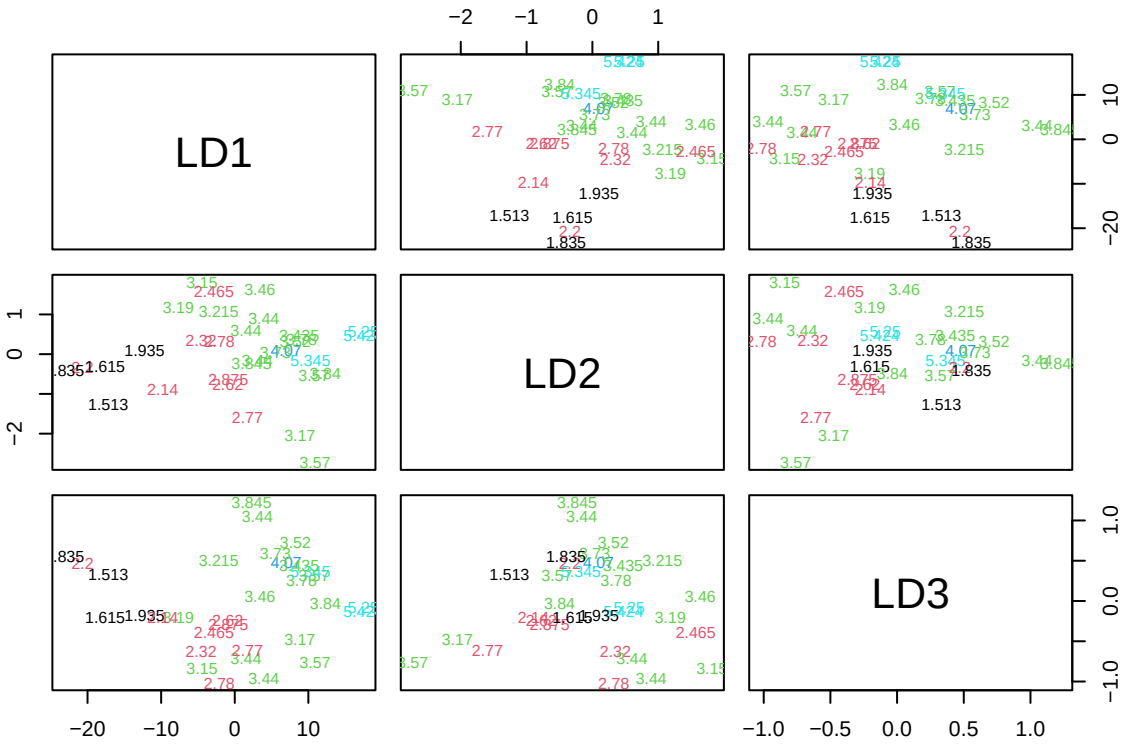
```
lda_model = lda(wt ~ ., data = mtcars)
```

```
## Warning in lda.default(x, grouping, ...): variables are collinear
```

```
print(lda_model)
```

```
## Call:
## lda(wt ~ ., data = mtcars)
##
## Prior probabilities of groups:
##   1.513   1.615   1.835   1.935    2.14     2.2    2.32   2.465    2.62    2.77
## 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125
##    2.78   2.875    3.15    3.17    3.19   3.215   3.435    3.44    3.46    3.52
## 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.09375 0.03125 0.03125
##    3.57    3.73    3.78    3.84   3.845    4.07    5.25   5.345   5.424
## 0.06250 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125 0.03125
##
## Group means:
##            mpg      cyl     disp       hp     drat     qsec        vs  am
## 1.513 30.40000 4.000000  95.1000 113.0000 3.770000 16.90000 1.0000000 1.0
## 1.615 30.40000 4.000000  75.7000  52.0000 4.930000 18.52000 1.0000000 1.0
## 1.835 33.90000 4.000000  71.1000  65.0000 4.220000 19.90000 1.0000000 1.0
## 1.935 27.30000 4.000000  79.0000  66.0000 4.080000 18.90000 1.0000000 1.0
## 2.14  26.00000 4.000000 120.3000  91.0000 4.430000 16.70000 0.0000000 1.0
## 2.2   32.40000 4.000000  78.7000  66.0000 4.080000 19.47000 1.0000000 1.0
## 2.32  22.80000 4.000000 108.0000  93.0000 3.850000 18.61000 1.0000000 1.0
## 2.465 21.50000 4.000000 120.1000  97.0000 3.700000 20.01000 1.0000000 0.0
## 2.62  21.00000 6.000000 160.0000 110.0000 3.900000 16.46000 0.0000000 1.0
## 2.77  19.70000 6.000000 145.0000 175.0000 3.620000 15.50000 0.0000000 1.0
## 2.78  21.40000 4.000000 121.0000 109.0000 4.110000 18.60000 1.0000000 1.0
## 2.875 21.00000 6.000000 160.0000 110.0000 3.900000 17.02000 0.0000000 1.0
## 3.15  22.80000 4.000000 140.8000  95.0000 3.920000 22.90000 1.0000000 0.0
## 3.17  15.80000 8.000000 351.0000 264.0000 4.220000 14.50000 0.0000000 1.0
## 3.19  24.40000 4.000000 146.7000  62.0000 3.690000 20.00000 1.0000000 0.0
## 3.215 21.40000 6.000000 258.0000 110.0000 3.080000 19.44000 1.0000000 0.0
## 3.435 15.20000 8.000000 304.0000 150.0000 3.150000 17.30000 0.0000000 0.0
## 3.44  18.56667 6.666667 231.7333 140.3333 3.663333 18.07333 0.6666667 0.0
## 3.46  18.10000 6.000000 225.0000 105.0000 2.760000 20.22000 1.0000000 0.0
## 3.52  15.50000 8.000000 318.0000 150.0000 2.760000 16.87000 0.0000000 0.0
## 3.57  14.65000 8.000000 330.5000 290.0000 3.375000 15.22000 0.0000000 0.5
## 3.73  17.30000 8.000000 275.8000 180.0000 3.070000 17.60000 0.0000000 0.0
## 3.78  15.20000 8.000000 275.8000 180.0000 3.070000 18.00000 0.0000000 0.0
## 3.84  13.30000 8.000000 350.0000 245.0000 3.730000 15.41000 0.0000000 0.0
## 3.845 19.20000 8.000000 400.0000 175.0000 3.080000 17.05000 0.0000000 0.0
## 4.07  16.40000 8.000000 275.8000 180.0000 3.070000 17.40000 0.0000000 0.0
## 5.25  10.40000 8.000000 472.0000 205.0000 2.930000 17.98000 0.0000000 0.0
## 5.345 14.70000 8.000000 440.0000 230.0000 3.230000 17.42000 0.0000000 0.0
```

```
## 5.424 10.40000 8.000000 460.0000 215.0000 3.000000 17.82000 0.0000000 0.0
##           gear     carb
## 1.513 5.000000 2.000000
## 1.615 4.000000 2.000000
## 1.835 4.000000 1.000000
## 1.935 4.000000 1.000000
## 2.14  5.000000 2.000000
## 2.2   4.000000 1.000000
## 2.32  4.000000 1.000000
## 2.465 3.000000 1.000000
## 2.62  4.000000 4.000000
## 2.77  5.000000 6.000000
## 2.78  4.000000 2.000000
## 2.875 4.000000 4.000000
## 3.15  4.000000 2.000000
## 3.17  5.000000 4.000000
## 3.19  4.000000 2.000000
## 3.215 3.000000 1.000000
## 3.435 3.000000 2.000000
## 3.44  3.666667 3.333333
## 3.46  3.000000 1.000000
## 3.52  3.000000 2.000000
## 3.57  4.000000 6.000000
## 3.73  3.000000 3.000000
## 3.78  3.000000 3.000000
## 3.84  3.000000 4.000000
## 3.845 3.000000 2.000000
## 4.07  3.000000 3.000000
## 5.25  3.000000 4.000000
## 5.345 3.000000 4.000000
## 5.424 3.000000 4.000000
##
## Coefficients of linear discriminants:
##                LD1           LD2           LD3
## mpg  -1.6686706267 -0.1115896854  0.1396977296
## cyl   0.1421739359 -0.1151987914  0.1588534606
## disp  0.0007828128 -0.0004250021  0.0018545594
## hp    0.0058000379 -0.0061406676 -0.0002669514
## drat -0.1281729007  0.0350876256 -0.4651870081
## qsec  0.1378721393  0.2224683387 -0.0995316577
## vs   -0.2843478718  0.2303975827 -0.3177069213
## am    0.5352954619 -0.6211592066 -0.2794715015
## gear  0.1296488303 -0.1753353068 -0.1842285434
## carb  0.0648244151 -0.0876676534 -0.0921142717
##
## Proportion of trace:
##    LD1    LD2    LD3
## 0.9896 0.0083 0.0021
plot(lda_model, col = as.numeric(mtcars$wt))
```




```
data(airquality)
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
aq <- na.omit(airquality)

aq$TempHigh <- ifelse(aq$Temp > 80, 1, 0)

table(aq$TempHigh)
```

```
##
##  0  1
## 60 51
```

```
model_logit <- glm(TempHigh ~ Solar.R + Wind, data = aq, family = binomial)

summary(model_logit)
```

```
##
## Call:
## glm(formula = TempHigh ~ Solar.R + Wind, family = binomial, data = aq)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.541919   0.856228   1.801 0.071730 .
```

```
## Solar.R      0.005142   0.002450   2.099 0.035848 *
## Wind        -0.274220   0.073534  -3.729 0.000192 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 153.15  on 110  degrees of freedom
## Residual deviance: 128.95  on 108  degrees of freedom
## AIC: 134.95
##
## Number of Fisher Scoring iterations: 4
```