STK2100: Solutions Week 15

Lars H. B. Olsen

20.04.2021

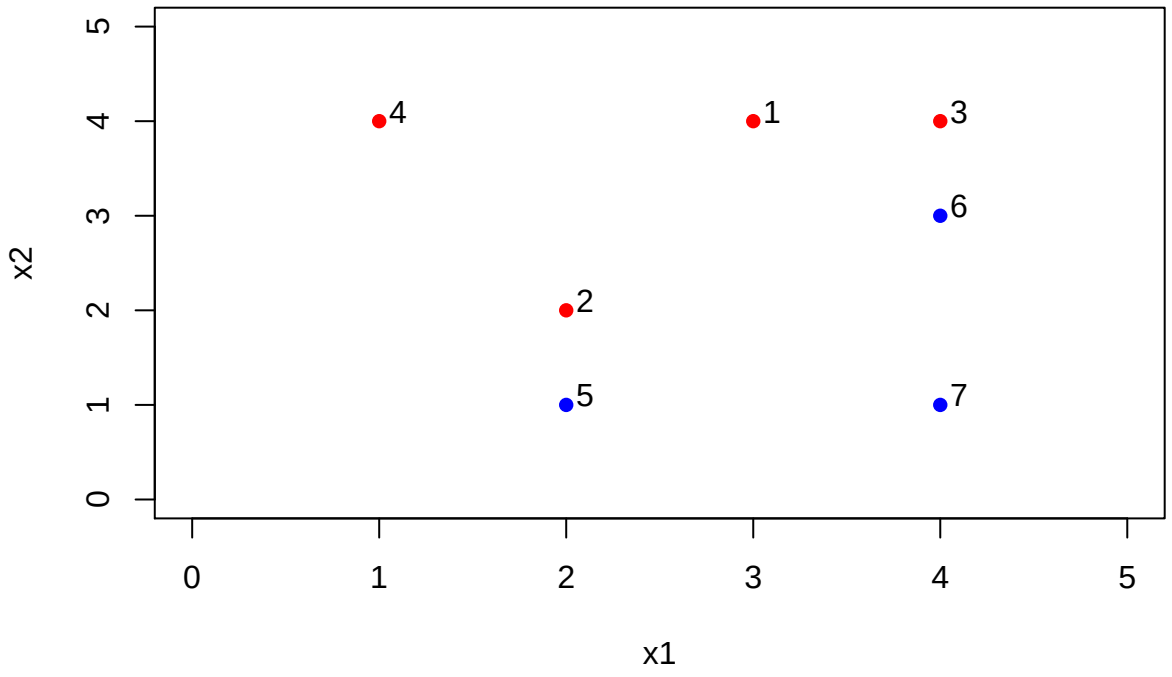
ISLR

Exercise 9.3

a)

Plot the given points, with corresponding color.

```
x1 = c(3, 2, 4, 1, 2, 4, 4)
x2 = c(4, 2, 4, 4, 1, 3, 1)
colors = c("red", "red", "red", "red", "blue", "blue", "blue")
plot(x1, x2, col = colors, xlim = c(0, 5), ylim = c(0, 5), pch = 16)
text(x1+0.1, x2+0.1, labels = seq(7))
```



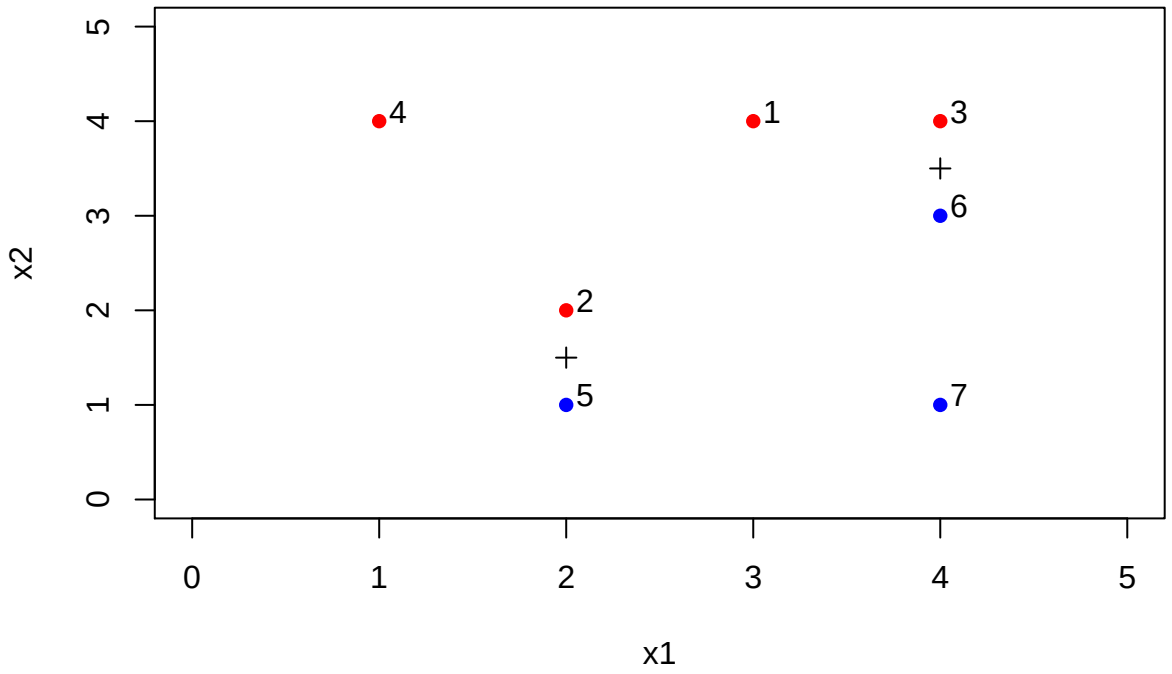
b)

The maximal margin classifier has to be in between observations #2, #3 and #5, #6. Locate the midpoints between #2 and #5, and #3 and 6#.

```
mid1 = c((x1[2] + x1[5])/2, (x2[2] + x2[5])/2)
mid2 = c((x1[3] + x1[6])/2, (x2[3] + x2[6])/2)
show(rbind(mid1, mid2))
```

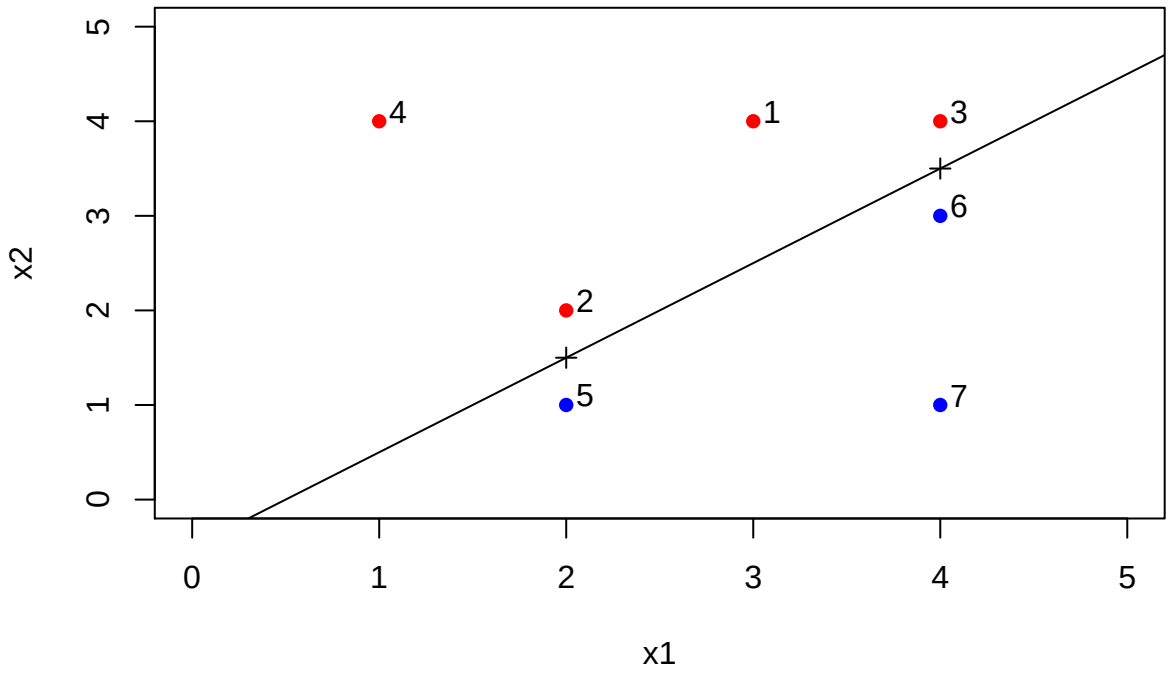
```
##      [,1] [,2]
## mid1    2  1.5
## mid2    4  3.5
```

```
plot(x1, x2, col = colors, xlim = c(0, 5), ylim = c(0, 5), pch = 16)
text(x1+0.1, x2+0.1, labels = seq(7))
points(rbind(mid1, mid2), col = 1, pch = 3)
```



They have cordinates $(2, 1.5)$ and $(4, 3.5)$, respectively. Find slope and intersect of the line that intersect these two points. We get slope $a = \frac{\Delta x_2}{\Delta x_1} = \frac{3.5-1.5}{4-2} = 1$ and intersect $b = x_2 - ax_1 = 1.5 - 2 = -0.5$. Plot the optimal seperating hyperplane $x_2 = ax_1 + b = x_1 - 0.5$.

```
plot(x1, x2, col = colors, xlim = c(0, 5), ylim = c(0, 5), pch = 16)
text(x1+0.1, x2+0.1, labels = seq(7))
points(rbind(mid1, mid2), col = 1, pch = 3)
abline(-0.5, 1)
```



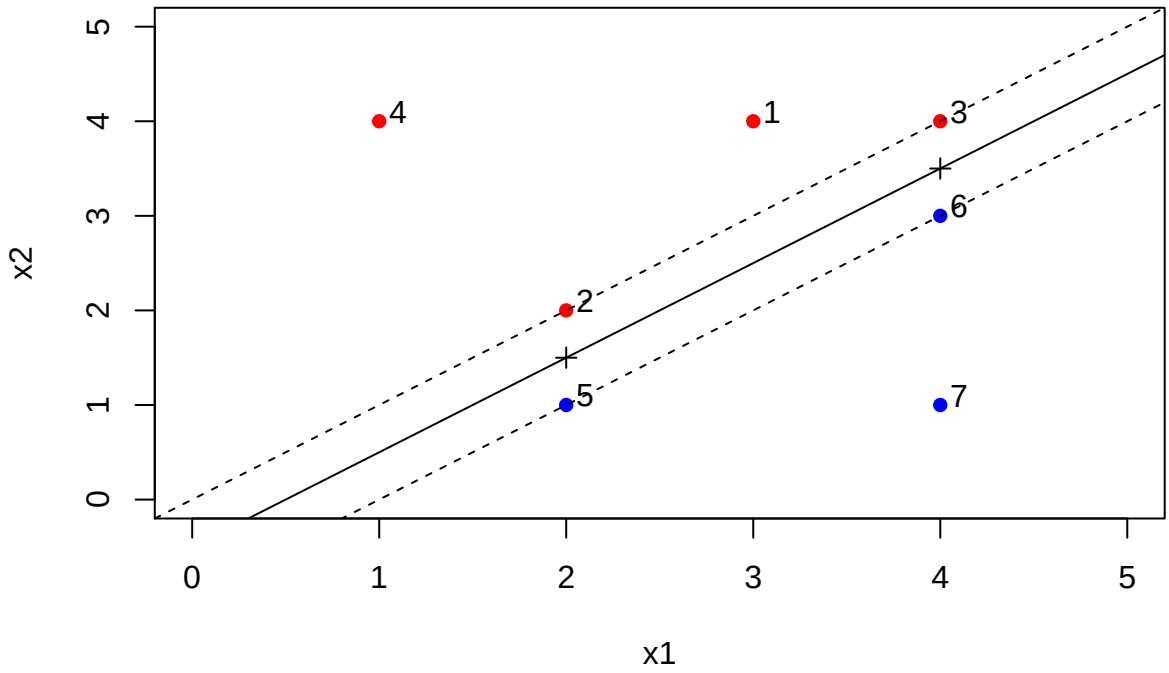
c)

Classify to red if we are above the line and blue if we are below. That is, for this dataset, we classify to red if $x_2 > ax_1 + b \Leftrightarrow x_2 > x_1 - 0.5 \Leftrightarrow 0.5 - x_1 + x_2 > 0$ and to blue otherwise.

d)

We can move the separation line up or down 0.5 units before we either hit a red or blue point, respectively. That is, we get the following two lines $l_1 = x_1 - 0.5 + 0.5 = x_1$ and $l_2 = x_1 - 0.5 - 0.5 = x_1 - 1$.

```
plot(x1, x2, col = colors, xlim = c(0, 5), ylim = c(0, 5), pch = 16)
text(x1+0.1, x2+0.1, labels = seq(7))
points(rbind(mid1, mid2), col = 1, pch = 3)
abline(-0.5, 1)
abline(-1, 1, lty = 2)
abline(0, 1, lty = 2)
```



The margin is the minimum **perpendicular** distance from the support vectors to the separation line. Use formula for distance from point to a line.

```
a = -1
b = 1
c = 0.5

margins = rep(NA, length(x1))
loc_x = rep(NA, length(x1))
loc_y = rep(NA, length(x1))

for (obs in seq(length(x1))) {
  margins[obs] = abs(a*x1[obs] + b*x2[obs] + c) / sqrt(a^2 + b^2)
  loc_x[obs] = (b*(b*x1[obs] - a*x2[obs]) - a*c) / (a^2 + b^2)
  loc_y[obs] = (a*(-b*x1[obs] + a*x2[obs]) - b*c) / (a^2 + b^2)
}
min(margins)
```

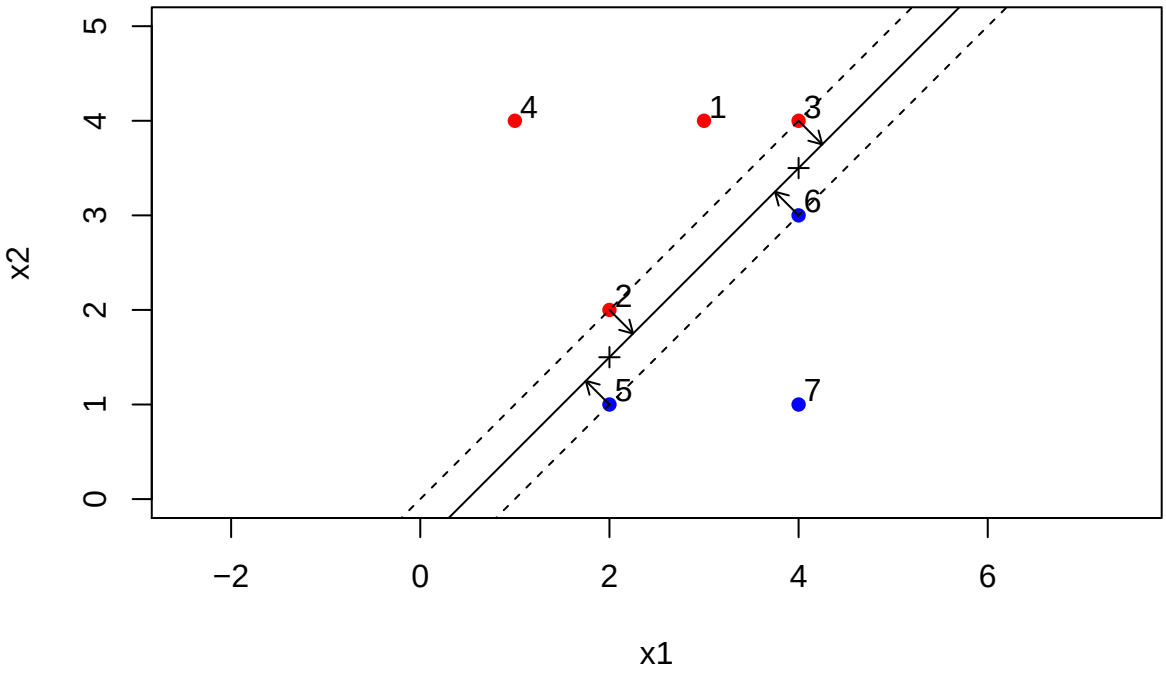
```
## [1] 0.3535534
```

We get a margin of 0.354.

Note that I set the aspect ration to 1, otherwise the perpendicular segments do not appear perpedicular, even though they are.

```
plot(x1, x2, col = colors, xlim = c(0, 5), ylim = c(0, 5), pch = 16, asp = 1)
text(x1+0.15, x2+0.15, labels = seq(7))
points(rbind(mid1, mid2), col = 1, pch = 3)
abline(-0.5, 1)
abline(-1, 1, lty = 2)
abline(0, 1, lty = 2)

arrows(x1[5], x2[5], loc_x[5], loc_y[5], length = 0.075)
arrows(x1[6], x2[6], loc_x[6], loc_y[6], length = 0.075)
arrows(x1[2], x2[2], loc_x[2], loc_y[2], length = 0.075)
```

```
arrows(x1[3], x2[3], loc_x[3], loc_y[3], length = 0.075)
```



e)

The support vectors are the coordinates of the observations that lie along the dashed lines. See page 341 and onward in the book. Thus, The support vectors are the points $(2, 1)$, $(2, 2)$, $(4, 3)$, and $(4, 4)$.

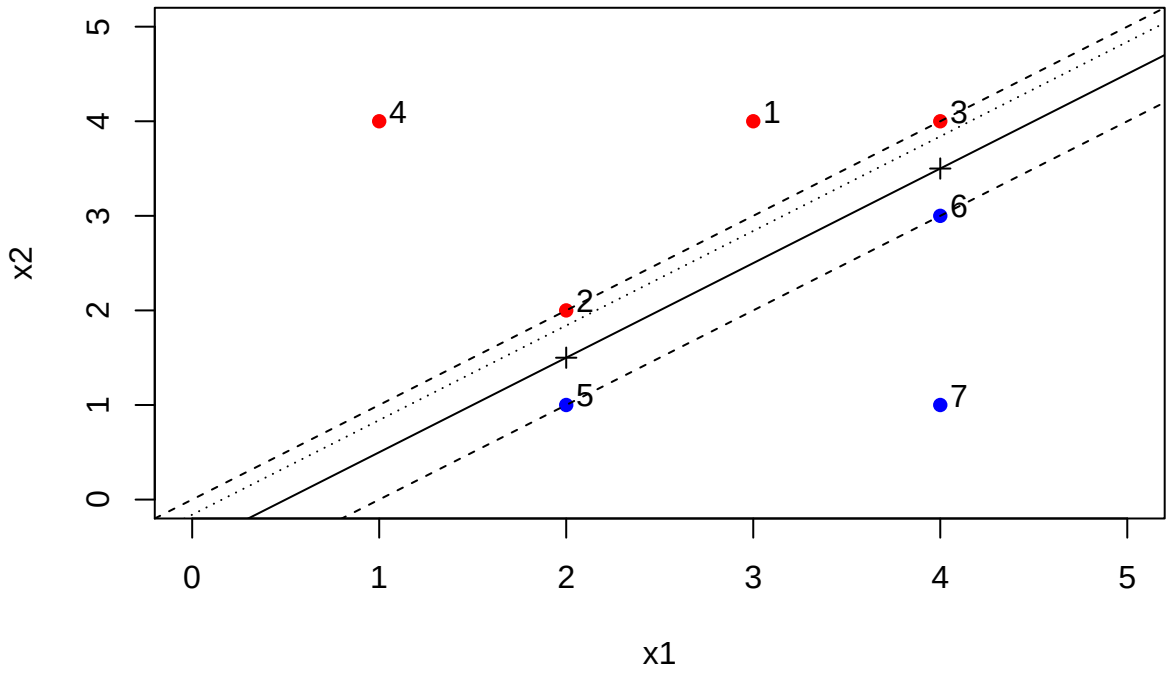
f)

By examining the plot, it is clear that if we moved the observation $(4, 1)$, we would not change the maximal margin hyperplane as it is not a support vector.

g)

Infinitely many solutions, can set intersect to be any value between -1 and 0, exclusively. I choose $x_2 = x_1 - 0.16$.

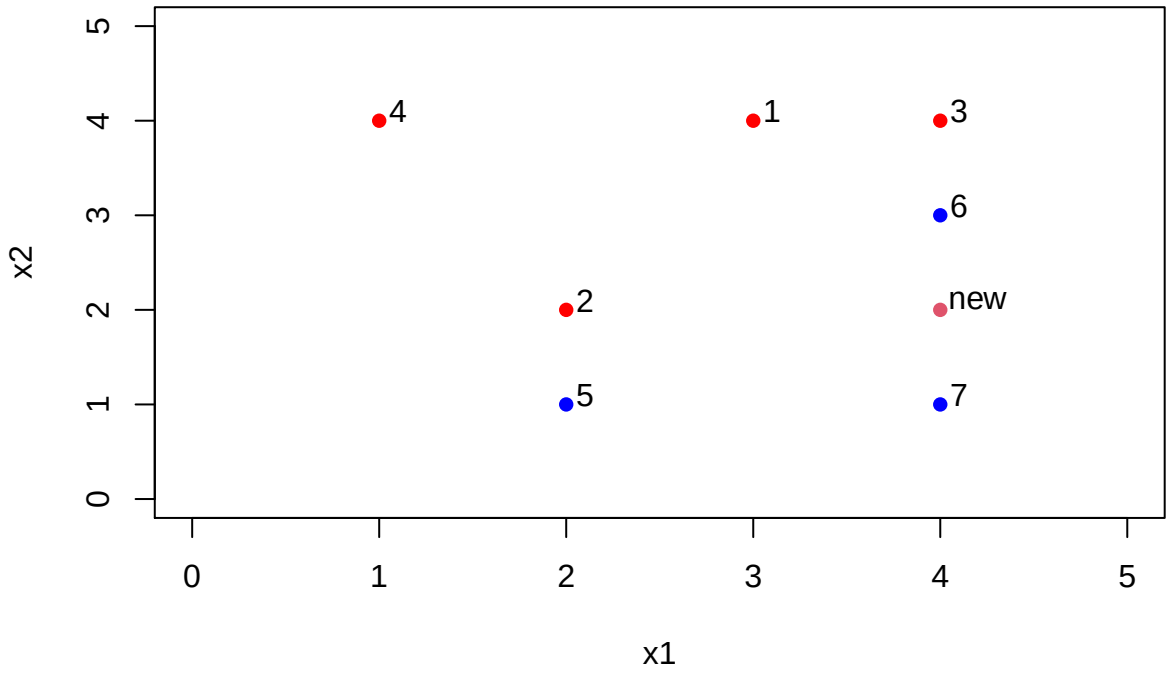
```
plot(x1, x2, col = colors, xlim = c(0, 5), ylim = c(0, 5), pch = 16)
text(x1+0.1, x2+0.1, labels = seq(7))
points(rbind(mid1, mid2), col = 1, pch = 3)
abline(-0.5, 1)
abline(-1, 1, lty = 2)
abline(0, 1, lty = 2)
abline(-0.16, 1, lty = 3)
```



h)

Once again, infinitely many solutions. For example by getting a red observation with coordinates $(4, 2)$. Then the two classes are obviously not separable by a hyperplane anymore.

```
plot(x1, x2, col = colors, xlim = c(0, 5), ylim = c(0, 5), pch = 16)
text(x1+0.1, x2+0.1, labels = seq(7))

points(4, 2, col = 2, pch = 16)
text(4+0.2, 2+0.1, labels = "new")
```



Exercise 9.7

a)

Create a binary variable that takes on a 1 for cars with gas mileage above the median, and a 0 for cars with gas mileage below the median.

```
library(ISLR)
var <- ifelse(Auto$mpg > median(Auto$mpg), 1, 0)
Auto$mpglevel <- as.factor(var)
```

b)

Fit a support vector classifier to the data with various values of "cost", in order to predict whether a car gets high of low gas mileage. Report the cross-validation errors associated with different values of this parameter. Comment on your results.

```
library(e1071)
set.seed(1)
tune.out <- tune(svm, mpglevel ~ ., data = Auto, kernel = "linear",
                 ranges = list(cost = c(0.01, 0.1, 1, 5, 10, 100, 1000)))
summary(tune.out)
```

```
##
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##  cost
##     1
##
## - best performance: 0.01025641
##
## - Detailed performance results:
##    cost      error dispersion
## 1 1e-02 0.07653846 0.03617137
## 2 1e-01 0.04596154 0.03378238
## 3 1e+00 0.01025641 0.01792836
## 4 5e+00 0.02051282 0.02648194
## 5 1e+01 0.02051282 0.02648194
## 6 1e+02 0.03076923 0.03151981
## 7 1e+03 0.03076923 0.03151981
```

A cost of 1 seems to perform best.

c)

Now repeat (b), this time using SVMs with radial and polynomial basis kernels, with different values of "gamma" and "degree" and "cost". Comment on your results.

```
set.seed(1)
tune.out <- tune(svm, mpglevel ~ ., data = Auto, kernel = "polynomial",
                 ranges = list(cost = c(0.01, 0.1, 1, 5, 10, 100),
                               degree = c(2, 3, 4)))
summary(tune.out)
```

```
##
```

```
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##  cost degree
##   100      2
##
## - best performance: 0.3013462
##
## - Detailed performance results:
##     cost degree     error dispersion
## 1  1e-02      2 0.5511538 0.04366593
## 2  1e-01      2 0.5511538 0.04366593
## 3  1e+00      2 0.5511538 0.04366593
## 4  5e+00      2 0.5511538 0.04366593
## 5  1e+01      2 0.5130128 0.08963366
## 6  1e+02      2 0.3013462 0.09961961
## 7  1e-02      3 0.5511538 0.04366593
## 8  1e-01      3 0.5511538 0.04366593
## 9  1e+00      3 0.5511538 0.04366593
## 10 5e+00      3 0.5511538 0.04366593
## 11 1e+01      3 0.5511538 0.04366593
## 12 1e+02      3 0.3446154 0.09821588
## 13 1e-02      4 0.5511538 0.04366593
## 14 1e-01      4 0.5511538 0.04366593
## 15 1e+00      4 0.5511538 0.04366593
## 16 5e+00      4 0.5511538 0.04366593
## 17 1e+01      4 0.5511538 0.04366593
## 18 1e+02      4 0.5511538 0.04366593
```

For a polynomial kernel, the lowest cross-validation error is obtained for a degree of 2 and a cost of 100.

```
set.seed(12)
tune.out <- tune(svm, mpglevel ~ ., data = Auto, kernel = "radial",
                 ranges = list(cost = c(0.01, 0.1, 1, 5, 10, 100),
                               gamma = c(0.01, 0.1, 1, 5, 10, 100)))
summary(tune.out)
```

```
##
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##  cost gamma
##   100  0.01
##
## - best performance: 0.01025641
##
## - Detailed performance results:
##     cost gamma      error dispersion
## 1  1e-02 1e-02 0.55365385 0.05054135
## 2  1e-01 1e-02 0.08660256 0.04333178
## 3  1e+00 1e-02 0.07141026 0.04286646
```

```
## 4  5e+00 1e-02 0.04583333 0.03745995
## 5  1e+01 1e-02 0.02551282 0.02689559
## 6  1e+02 1e-02 0.01025641 0.01324097
## 7  1e-02 1e-01 0.16596154 0.10427983
## 8  1e-01 1e-01 0.07903846 0.04234873
## 9  1e+00 1e-01 0.05096154 0.03789619
## 10 5e+00 1e-01 0.02557692 0.02702877
## 11 1e+01 1e-01 0.02301282 0.02549182
## 12 1e+02 1e-01 0.02814103 0.02822527
## 13 1e-02 1e+00 0.55365385 0.05054135
## 14 1e-01 1e+00 0.55365385 0.05054135
## 15 1e+00 1e+00 0.05858974 0.03963453
## 16 5e+00 1e+00 0.05865385 0.03806449
## 17 1e+01 1e+00 0.05865385 0.03806449
## 18 1e+02 1e+00 0.05865385 0.03806449
## 19 1e-02 5e+00 0.55365385 0.05054135
## 20 1e-01 5e+00 0.55365385 0.05054135
## 21 1e+00 5e+00 0.49750000 0.06178906
## 22 5e+00 5e+00 0.49493590 0.05667386
## 23 1e+01 5e+00 0.49493590 0.05667386
## 24 1e+02 5e+00 0.49493590 0.05667386
## 25 1e-02 1e+01 0.55365385 0.05054135
## 26 1e-01 1e+01 0.55365385 0.05054135
## 27 1e+00 1e+01 0.50269231 0.06670076
## 28 5e+00 1e+01 0.50012821 0.06676097
## 29 1e+01 1e+01 0.50012821 0.06676097
## 30 1e+02 1e+01 0.50012821 0.06676097
## 31 1e-02 1e+02 0.55365385 0.05054135
## 32 1e-01 1e+02 0.55365385 0.05054135
## 33 1e+00 1e+02 0.55365385 0.05054135
## 34 5e+00 1e+02 0.55365385 0.05054135
## 35 1e+01 1e+02 0.55365385 0.05054135
## 36 1e+02 1e+02 0.55365385 0.05054135
```

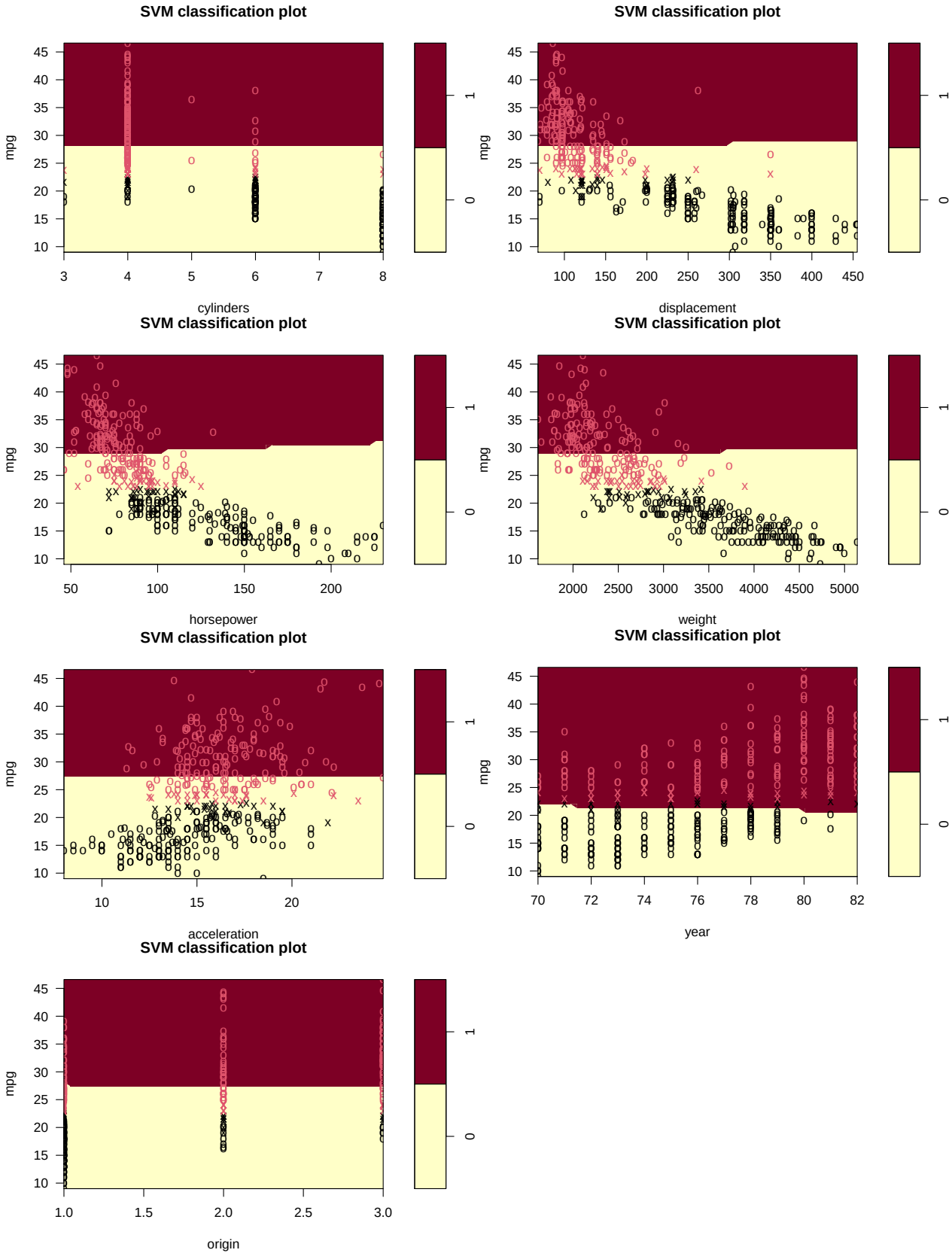
For a radial kernel, the lowest cross-validation error is obtained for a gamma of 0.01 and a cost of 100.

d)

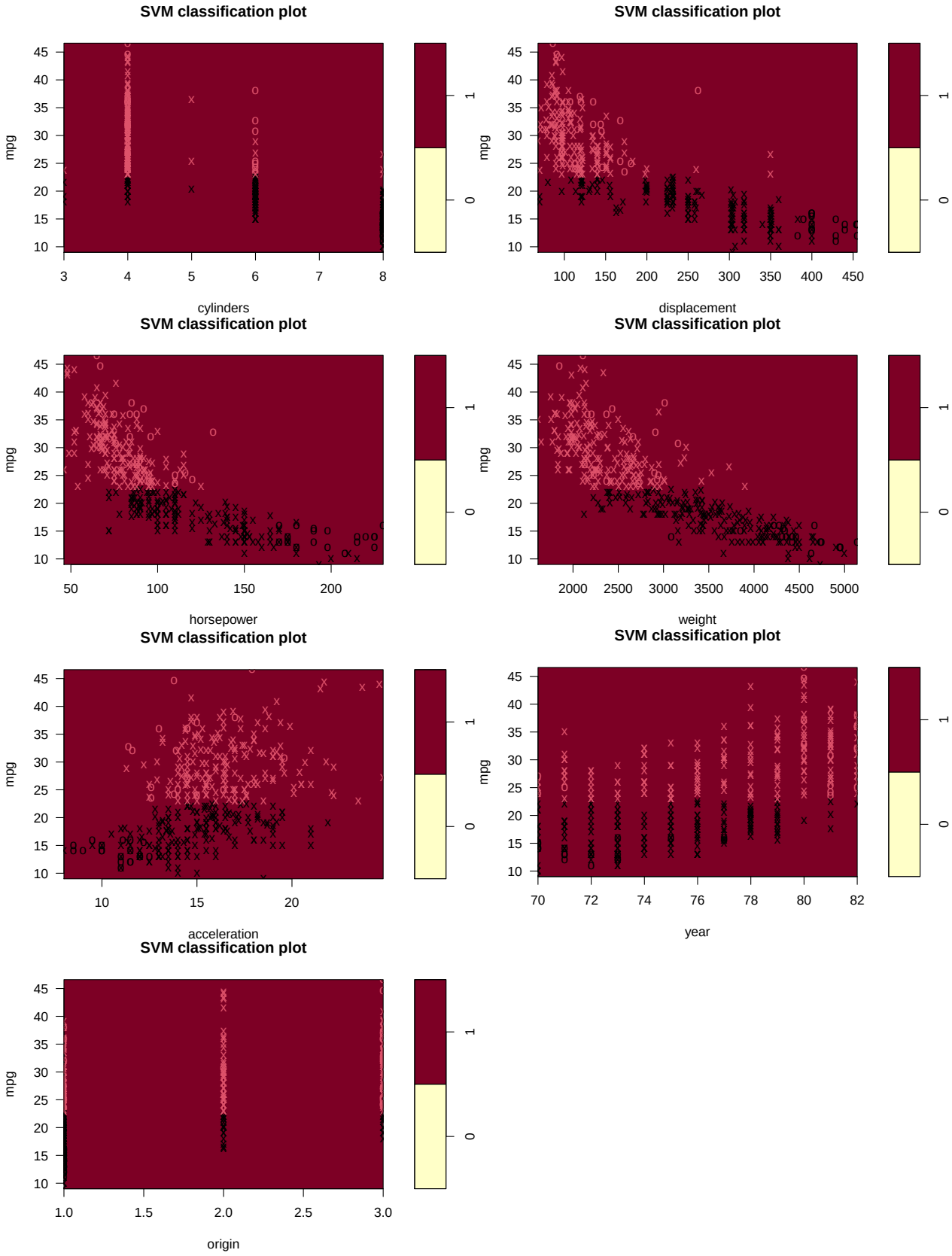
Make some plots to back up your assertions in (b) and (c). We saw that the linear kernel performed best. Thus, we expect the corresponding separations to be good, while the other will perform worse. Which we also see by the pictures below.

```r
# Fit the optimal models according to the results above
svm.linear <- svm(mpglevel ~ ., data = Auto, kernel = "linear", cost = 1)
svm.poly <- svm(mpglevel ~ ., data = Auto, kernel = "polynomial", cost = 100, degree = 2)
svm.radial <- svm(mpglevel ~ ., data = Auto, kernel = "radial", cost = 100, gamma = 0.01)

# Function that create the plots of the applicable variables
plotpairs = function(fit) {
    for (name in names(Auto)[!(names(Auto) %in% c("mpg", "mpglevel", "name"))]) {
        plot(fit, Auto, as.formula(paste("mpg~", name, sep = "")))
    }
}

# Plot the functions
```

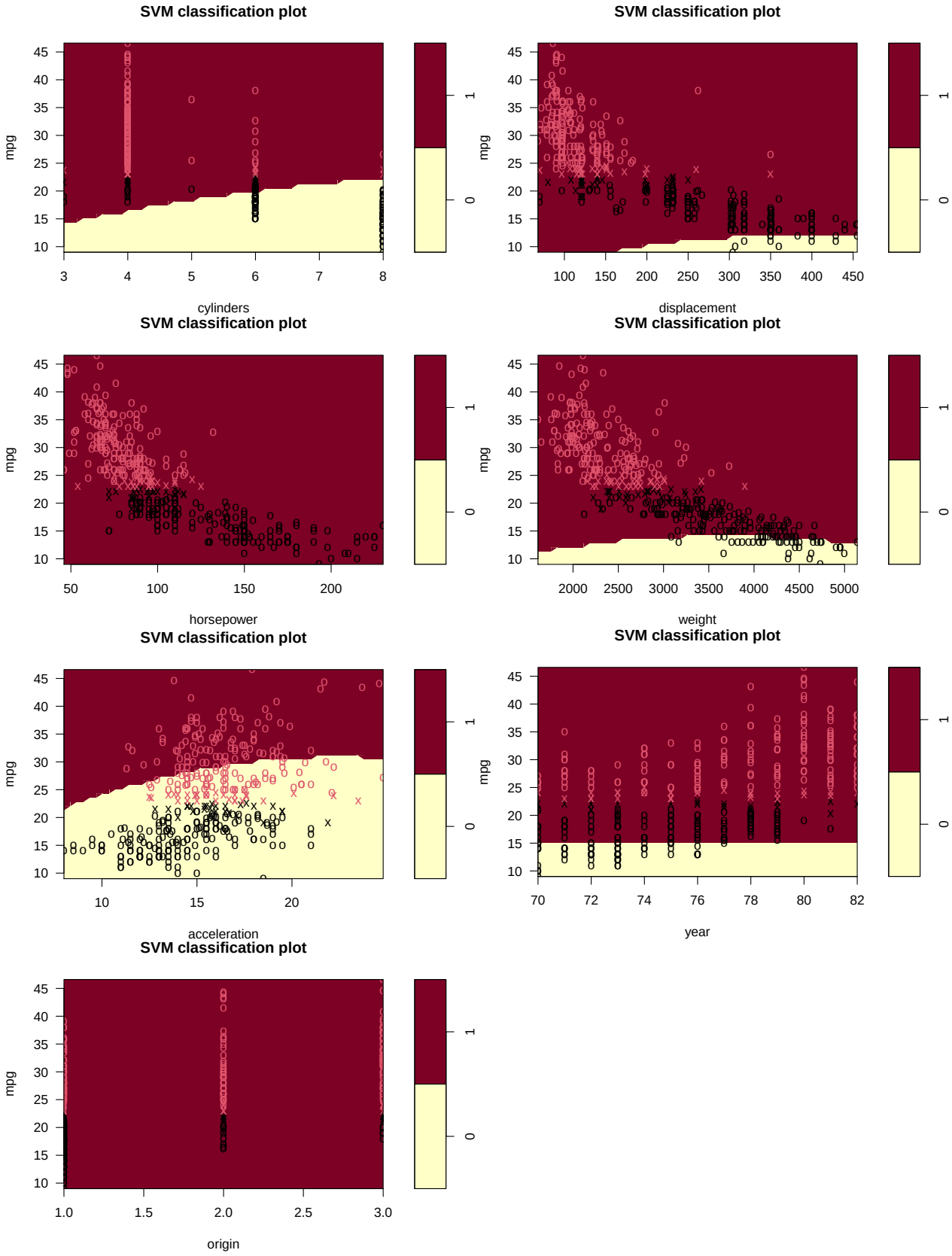
```
plotpairs(svm.linear)
```



```
plotpairs(svm.poly)
```



```
plotpairs(svm.radial)
```



Extra

Exercise 6

Rewrite the loss function as $L(y, f(x)) = c_0 I\{y = 0,\, f(x) = 1\} + c_1 I\{y = 1,\, f(x) = 0\}$. Let $Q_0(x) = c_0 \Pr(Y = 0|X = x)$ and $Q_1(x) = c_1 \Pr(Y = 1|X = x)$. We get that

$$
\begin{aligned}
\mathrm{EPE}(f) &= \mathrm{E}[L(y, f(x))] \\
&= \int_x \int_y \left(c_0 I\{y = 0,\, f(x) = 1\} + c_1 I\{y = 1,\, f(x) = 0\}\right) p(y|x)dy\, p(x)dx \\
&= \int_x \int_y c_0 I\{y = 0,\, f(x) = 1\} p(y|x)dy\, p(x)dx + \int_x \int_y c_1 I\{y = 1,\, f(x) = 0\} p(y|x)dy\, p(x)dx \\
&= \int_{x:f(x)=1} \int_y c_0 I\{y = 0\} p(y|x)dy\, p(x)dx + \int_{x:f(x)=0} \int_y c_1 I\{y = 1\} p(y|x)dy\, p(x)dx \\
&= \int_{x:f(x)=1} c_0 \Pr(Y = 0|X = x) p(x)dx + \int_{x:f(x)=0} c_1 p(Y = 1|X = x) p(x)dx \\
&= \int_{x:f(x)=1} Q_0(x) p(x)dx + \int_{x:f(x)=0} Q_1(x) p(x)dx \\
&= \int_x I\{f(x) = 1\} Q_0(x) p(x)dx + \int_x I\{f(x) = 0\} Q_1(x) p(x)dx \\
&= \int_x \left(I\{f(x) = 1\} Q_0(x) + I\{f(x) = 0\} Q_1(x)\right) p(x)dx \\
&= \int_x \left(I\{f(x) = 1\} Q_0(x) + I\{f(x) = 0\} Q_1(x) - I\{f(x) = 0\} Q_0(x) + I\{f(x) = 0\} Q_0(x)\right) p(x)dx \\
&= \int_x \left(I\{f(x) = 0\} \left[Q_1(x) - Q_0(x)\right] + Q_0(x) \left[I\{f(x) = 0\} + I\{f(x) = 1\}\right]\right) p(x)dx \\
&= \int_x Q_0(x) p(x)dx + \int_x I\{f(x) = 0\} \left[Q_1(x) - Q_0(x)\right] p(x)dx \\
&= \mathrm{Const} + \int_x I\{f(x) = 0\} \left[Q_1(x) - Q_0(x)\right] p(x)dx
\end{aligned}
$$

We want to minimize $\mathrm{EPE}(f)$, that is the same as minimzing $I\{f(x) = 0\} [Q_1(x) - Q_0(x)]$. Recall that we can only alter this expression by changing f. So if $Q_1(x) - Q_0(x)$ is positive, we should set $f(x) = 1$, as otherwise this term will contribute positively to $\mathrm{EPE}(f)$. Similarily, when $Q_1(x) - Q_0(x)$ is negative, we would set $f(x) = 0$, as this will reduce the $\mathrm{EPE}(f)$.

In other words, we set $f(x) = 1$ if $Q_1(x) - Q_0(x) > 0 \Leftrightarrow c_1 \Pr(Y = 1|X = x) > c_0 \Pr(Y = 0|X = x) \Leftrightarrow \Pr(Y = 1|X = x) > \frac{c_0}{c_1} \Pr(Y = 0|X = x)$, and $f(x) = 0$ otherwise.

This is a reasonable classification rule. If $c_0 = c_1$, we get the version from Exercise 3. Furthermore, if, e.g., $c_0 = 2c_1$, then $\Pr(Y = 1|X = x)$ has to be twice as big as $\Pr(Y = 0|X = x)$ for us to set $f(x) = 1$. Reasonable, as the related loss of false positives are twice as large as for false negatives.

Exercise 7

Solutions are provided by Vinnie Ko.

(a)

(i)

$p = 0$, so the model: $Y_i = \beta_0 + \varepsilon_i$, and $X = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

The least squares estimate is given by:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y},$$

which in this case leads to

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$$

So,

$$\widehat{y}_i = \overline{y} \text{ for } 1 \leq i \leq n.$$

(ii)

Same procedure as in (i), but you have to replace $\boldsymbol{X}$ and $\boldsymbol{y}$ with $\boldsymbol{X}_{-i}$ and $\boldsymbol{y}_{-i}$ by removing the i-th data point.
The resulting prediction:

$$\widehat{y_i}^{-i} = \frac{1}{n-1}\sum_{j \neq i} y_i.$$

(iii)

$$\begin{aligned}
\boldsymbol{H} &= \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}} \\
&= \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \left(\begin{bmatrix} 1 \cdots 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 1 \cdots 1 \end{bmatrix} \\
&= n^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \cdots 1 \end{bmatrix} \\
&= \frac{1}{n} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}
\end{aligned}$$

Thus,

$$h_{ii} = \frac{1}{n}.$$

(iv)

$$
\begin{aligned}
y_i - \widehat{y}_i^{-i} &= y_i - \frac{\sum_{j \neq i} y_i}{n-1} \\
&= y_i - \frac{\sum_{i'=1}^{n} y_{i'} - y_i}{n-1} \\
&= y_i - \frac{\frac{\sum_{i'=1}^{n} y_{i'}}{n} - \frac{y_i}{n}}{\frac{n-1}{n}} \\
&= y_i - \frac{\widehat{y}_i - \frac{y_i}{n}}{1 - \frac{1}{n}} \\
&= \frac{(1 - \frac{1}{n}) y_i + \frac{y_i}{n} - \widehat{y}_i}{1 - \frac{1}{n}} \\
&= \frac{y_i - \widehat{y}_i}{1 - \frac{1}{n}} \\
&= \frac{y_i - \widehat{y}_i}{1 - h_i} && \text{(by using the result from (iii))}
\end{aligned}
$$

(b)

(i)

$$
\begin{aligned}
\boldsymbol{M}_n &= \boldsymbol{X}_n^{\mathrm{T}} \boldsymbol{X}_n \\
&= \begin{bmatrix} x_{1,1} & \cdots & x_{i,1} & \cdots & x_{n,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1,j} & \cdots & x_{i,j} & \cdots & x_{n,j} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{1,p} & \cdots & x_{i,p} & \cdots & x_{n,p} \end{bmatrix} \cdot \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,j} & \cdots & x_{n,p} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{x}_1 \cdots \boldsymbol{x}_n \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{bmatrix} \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}
\end{aligned}
$$

(ii)

$$
\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \right)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}
$$

if and only if

$$
\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \right) \left(\boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}} \right) = \boldsymbol{I} \text{ and } \left(\boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}} \right) \left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \right) = \boldsymbol{I}
$$

For convenience, lets write $c = \dfrac{1}{1 + \boldsymbol{v}^{\mathrm{T}} \boldsymbol{A}^{-1} \boldsymbol{u}}$.

First condition:

$$\begin{aligned}
\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right)\left(\boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}\right) &= \left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right)\left(\boldsymbol{A}^{-1} - c\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\right) \\
&= \boldsymbol{A}\boldsymbol{A}^{-1} - c\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} - c\boldsymbol{u}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u})\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} \\
&= \boldsymbol{I} - c\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} - c(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u})\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} \\
&= \boldsymbol{I} + (-c + 1 - c\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u})\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} \\
&= \boldsymbol{I} + (\frac{-1 + 1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u} - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}})\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} \\
&= \boldsymbol{I}
\end{aligned}$$

Second condition:

$$\begin{aligned}
\left(\boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}\right)\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right) &= \left(\boldsymbol{A}^{-1} - c\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\right)\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right) \\
&= \boldsymbol{A}^{-1}\boldsymbol{A} + \boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} - c\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} - c\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \\
&= \boldsymbol{I} + \left(1 - c - c\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right)\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \\
&= \boldsymbol{I} + (\frac{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u} - 1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}})\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \\
&= \boldsymbol{I}
\end{aligned}$$

Thus,

$$\left(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}.$$

(iii)

Let $\widetilde{\boldsymbol{x}}_n = \boldsymbol{M}_{n-1}^{-1}\boldsymbol{x}_n$, then

$$\begin{aligned}
\boldsymbol{M}_n^{-1} &= \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}\right)^{-1} \\
&= \left(\sum_{i=1}^{n-1} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} + \boldsymbol{x}_n \boldsymbol{x}_n^{\mathrm{T}}\right)^{-1} \\
&= \left(\boldsymbol{M}_{n-1}^{-1} + \boldsymbol{x}_n \boldsymbol{x}_n^{\mathrm{T}}\right)^{-1} \\
&= \boldsymbol{M}_{n-1}^{-1} - \frac{\boldsymbol{M}_{n-1}^{-1}\boldsymbol{x}_n \boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1 + \boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}\boldsymbol{x}_n} \qquad \text{(by using the Sherman-Morrison formula)} \\
&= \boldsymbol{M}_{n-1}^{-1} - \frac{\widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1 + \boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}
\end{aligned}$$

(iv)

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \widehat{\boldsymbol{\beta}}_n \\
&= (\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{y}_n \\
&= \boldsymbol{M}_n^{-1}\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{y}_n \\
&= \left(\boldsymbol{M}_{n-1}^{-1} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\sum_{i=1}^{n}\boldsymbol{x}_i y_i \\
&= \left(\boldsymbol{M}_{n-1}^{-1} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\left(\sum_{i=1}^{n-1}\boldsymbol{x}_i y_i + \boldsymbol{x}_n y_n\right) \\
&= \left(\boldsymbol{M}_{n-1}^{-1} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\left(\boldsymbol{X}_{n-1}^{\mathrm{T}}\boldsymbol{y}_{n-1} + \boldsymbol{x}_n y_n\right) \\
&= \boldsymbol{M}_{n-1}^{-1}\boldsymbol{X}_{n-1}^{\mathrm{T}}\boldsymbol{y}_{n-1} + \boldsymbol{M}_{n-1}^{-1}\boldsymbol{x}_n y_n - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\boldsymbol{X}_{n-1}^{\mathrm{T}}\boldsymbol{y}_{n-1} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\boldsymbol{x}_n y_n \\
&= \widehat{\boldsymbol{\beta}}_{n-1} + \boldsymbol{M}_{n-1}^{-1}\boldsymbol{x}_n y_n - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\widehat{\boldsymbol{\beta}}_{n-1} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{M}_{n-1}^{-1}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\boldsymbol{x}_n y_n \\
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widehat{\boldsymbol{\beta}}_{n-1} + \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\boldsymbol{M}_{n-1}^{-1}\boldsymbol{x}_n y_n \\
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widehat{\boldsymbol{\beta}}_{n-1} + \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widetilde{\boldsymbol{x}}_n y_n \\
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\left(\widehat{\boldsymbol{\beta}}_{n-1} + \widetilde{\boldsymbol{x}}_n y_n\right)
\end{aligned}
$$

Or alternatively,

$$
\begin{aligned}
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widehat{\boldsymbol{\beta}}_{n-1} + \widetilde{\boldsymbol{x}}_n y_n - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n y_n}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n} \\
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widehat{\boldsymbol{\beta}}_{n-1} + \widetilde{\boldsymbol{x}}_n y_n - \frac{\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n\widetilde{\boldsymbol{x}}_n y_n}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n} \\
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widehat{\boldsymbol{\beta}}_{n-1} + \left(1 - \frac{\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widetilde{\boldsymbol{x}}_n y_n \\
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widehat{\boldsymbol{\beta}}_{n-1} + \frac{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n - \boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\widetilde{\boldsymbol{x}}_n y_n \\
&= \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\widehat{\boldsymbol{\beta}}_{n-1} + \frac{1}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\widetilde{\boldsymbol{x}}_n y_n
\end{aligned}
$$

(v)

To prove $\left(\boldsymbol{I} - \dfrac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)^{-1} = \boldsymbol{I} + \widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}$, show

$$
\left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right)\left(\boldsymbol{I} + \widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\right) = \boldsymbol{I} \quad \text{and} \quad \left(\boldsymbol{I} + \widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}\right)\left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n\boldsymbol{x}_n^{\mathrm{T}}}{1+\boldsymbol{x}_n^{\mathrm{T}}\widetilde{\boldsymbol{x}}_n}\right) = \boldsymbol{I}
$$

or simply use the Sherman-Morison formula.

Use the result from (iv):

$$\widehat{\boldsymbol{\beta}}_n = \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}}}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}\right)\left(\widehat{\boldsymbol{\beta}}_{n-1} + \widetilde{\boldsymbol{x}}_n y_n\right)$$

$$\left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}}}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}\right)^{-1} \widehat{\boldsymbol{\beta}}_n = \widehat{\boldsymbol{\beta}}_{n-1} + \widetilde{\boldsymbol{x}}_n y_n$$

$$\widehat{\boldsymbol{\beta}}_{n-1} = \left(\boldsymbol{I} - \frac{\widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}}}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}\right)^{-1} \widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{x}}_n y_n.$$

Use the equality that we just proved

$$\widehat{\boldsymbol{\beta}}_{n-1} = \left(\boldsymbol{I} + \widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}}\right) \widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{x}}_n y_n.$$

(vi)

$$\begin{aligned}
y_n - \widehat{y}_n^{-n} &= y_n - \boldsymbol{x}_n^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{n-1} \\
&= y_n - \boldsymbol{x}_n^{\mathrm{T}} \left(\left(\boldsymbol{I} + \widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}}\right) \widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{x}}_n y_n\right) \\
&= y_n - \left(\left(\boldsymbol{x}_n^{\mathrm{T}} + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}}\right) \widehat{\boldsymbol{\beta}}_n - \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n y_n\right) \\
&= y_n - \left(\left(1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n\right) \boldsymbol{x}_n^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n - \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n y_n\right) \\
&= y_n + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n y_n - \left(1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n\right) \boldsymbol{x}_n^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n \\
&= \left(1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n\right) y_n - \left(1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n\right) \widehat{y}_n \\
&= \left(1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n\right) \left(y_n - \widehat{y}_n\right)
\end{aligned}$$

(vii)

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_n^{\mathrm{T}} \end{bmatrix} \cdot \boldsymbol{M}_n^{-1} \cdot \begin{bmatrix} \boldsymbol{x}_1 \cdots \boldsymbol{x}_n \end{bmatrix}$$

From this, we can directly see that $(\boldsymbol{H})_{i,j} = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{M}_n^{-1} \boldsymbol{x}_j$.

$$
\begin{aligned}
(\boldsymbol{H})_{n,n} &= \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{M}_n^{-1} \boldsymbol{x}_n \\
&= \boldsymbol{x}_n^{\mathrm{T}} \left(\boldsymbol{M}_{n-1}^{-1} - \frac{\widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{M}_{n-1}^{-1}}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n} \right) \boldsymbol{x}_n \\
&= \boldsymbol{x}_n^{\mathrm{T}} \left(\boldsymbol{M}_{n-1}^{-1} \boldsymbol{x}_n - \frac{\widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}} \boldsymbol{M}_{n-1}^{-1} \boldsymbol{x}_n}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n} \right) \\
&= \boldsymbol{x}_n^{\mathrm{T}} \left(\widetilde{\boldsymbol{x}}_n - \frac{\widetilde{\boldsymbol{x}}_n \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n} \right) \\
&= \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n \left(1 - \frac{\boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n} \right) \\
&= \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n \left(\frac{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n - \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n} \right) \\
&= \frac{\boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}
\end{aligned}
$$

First, we use the result from (vii):

$$
\begin{aligned}
h_n &= \frac{\boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n}{1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n} \\
h_n + h_n \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n &= \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n \\
h_n &= \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n - h_n \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n \\
\frac{h_n}{1 - h_n} &= \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n.
\end{aligned}
$$

We plug this result into the equation we just obtained:

$$
\begin{aligned}
y_n - \widehat{y}_n^{-n} &= \left(1 + \boldsymbol{x}_n^{\mathrm{T}} \widetilde{\boldsymbol{x}}_n\right) \left(y_n - \widehat{y}_n\right) \\
&= \left(1 + \frac{h_n}{1 - h_n}\right) \left(y_n - \widehat{y}_n\right) \\
&= \frac{y_n - \widehat{y}_n}{1 - h_n}.
\end{aligned}
$$

This verifies equation (5.2) in the textbook for $i = n$.

(viii)

Changing the order of data points in the dataset doesn't effect the model. This means that we can set any data point to be $\boldsymbol{x}_n$. Therefore, equation (5.2) is valid for all $i = 1, \cdots, n$.

(c)

Consider a situation where we fitted a model based on n data points. (i.e. We estimated $\widehat{\beta}_n$.) If we get m extra data points (after we already estimated $\widehat{\beta}_n$), we don't have to fit the whole model again, but we can just 'update' our model by using the formulas we obtained. (i.e. We can update $\widehat{\beta}_n$ to $\widehat{\beta}_{n+m}$.)