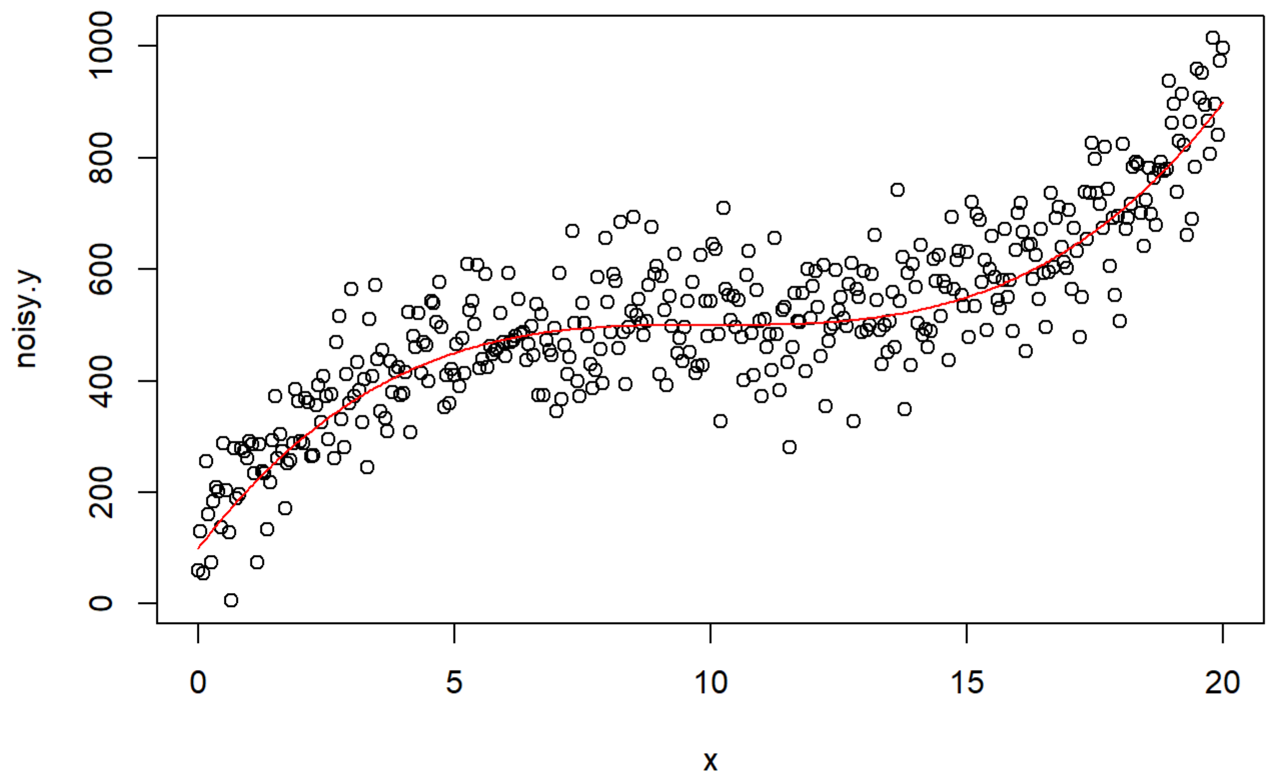
title: “Problem Set 2: Bias, Variance, Cross-Validation” author: “51709” date: | | 01 March 2023 output: pdf_document

1a. With predictor `x` and outcome `noisy_y`, split the data into a training and test set

```
library(boot)
set.seed(1)
x <- seq(from=0, to=20, by=0.05)
y <- 500 + 0.4 * (x-10)^3
noise <- rnorm(length(x), mean=10, sd=80)
noisy.y <- y + noise
{
plot(x,noisy.y)
lines(x, y, col='red')
}
```



Assign the data to a dataframe. Randomly sample the data, assigning 80% to training set and 20% to test set.

```
df <- data.frame(x, noisy.y)
N <- floor(.8*nrow(df))
train_idx <- sample(1:nrow(df), N)
df_train <- df[train_idx,]
df_test <- df[-train_idx,]
```

1b. Perform 10-fold CV for polynomials from degree 1 to 5 (use MSE as your error measure). This should be done from scratch using a for loop.

Randomly permute the training set and split into 10 evenly sized folds. Then create a for loop: for every polynomial model of degree 1 through 5, split the training set again into a training set and test set, with 1 fold in the test set and the rest in the training set. Change which fold is in the test set each time by inserting a nested for loop for i (fold used for test) 1 through 10 (for 10-fold cross-validation).

Calculate the mean-squared error of the test sub-set (validation set) for each fold. Store results from each loop in a results table with column headings for the degree of the model, the fold number and the mean-squared error.

```
folds <- sample(rep(1:10, each = nrow(df_train)/10), replace=FALSE)
degree <- 1:5
results <- data.frame(results_degree=numeric(0), results_i=numeric(0), results_error=numeric
(0))
for (d in degree){
  for (i in 1:10){
    cv.te <- df_train[folds==i,]
    cv.tr <- df_train[folds!=i,]
    cv.tr.x <- cv.tr[,1]
    cv.tr.y <- cv.tr[,2]
    cv.te.x <- cv.te[,1]
    cv.te.y <- cv.te[,2]
    mod <- glm(noisy.y ~ poly(x, d, raw = T), data=df_train)
    row <- data.frame(results_degree=d, results_i=i,     results_error=mean((predict(mod, cv.
te) - cv.te$noisy.y)^2))
    results <- rbind(results, row)
    }
}
results
```

```
##    results_degree results_i results_error
## 1               1         1      6420.144
## 2               1         2     10231.589
## 3               1         3      8808.168
## 4               1         4     12982.480
## 5               1         5     10610.066
## 6               1         6      8595.007
## 7               1         7      5252.368
## 8               1         8      6735.810
## 9               1         9      8653.017
## 10              1        10     13536.145
## 11              2         1      6421.101
## 12              2         2     10225.078
## 13              2         3      8810.706
## 14              2         4     12985.486
## 15              2         5     10619.164
## 16              2         6      8590.346
## 17              2         7      5251.758
## 18              2         8      6728.594
## 19              2         9      8656.569
## 20              2        10     13535.819
## 21              3         1      6285.938
## 22              3         2      6064.201
## 23              3         3      5801.355
## 24              3         4      8317.196
## 25              3         5      6565.344
## 26              3         6      4450.438
## 27              3         7      4031.181
## 28              3         8      4730.678
## 29              3         9      5092.755
## 30              3        10      9157.154
## 31              4         1      6279.820
## 32              4         2      6069.129
## 33              4         3      5813.017
## 34              4         4      8311.775
## 35              4         5      6572.819
## 36              4         6      4446.706
## 37              4         7      4025.874
## 38              4         8      4731.683
## 39              4         9      5088.472
## 40              4        10      9155.785
## 41              5         1      6218.422
## 42              5         2      6214.448
## 43              5         3      5759.784
## 44              5         4      8135.461
## 45              5         5      6536.950
## 46              5         6      4413.713
## 47              5         7      3970.983
## 48              5         8      4600.303
## 49              5         9      5245.364
## 50              5        10      9121.421
```

To calculate the overall cross-validation error for each model, create a table which summarises the cross-validation error for each fold. For each degree 1 through 5, calculate the sum of the MSE's across the 10 folds (`a`) using an ifelse condition, and calculate the number of folds (`b`). Then divide `a` by `b` to give the average

MSE for each of the 5 models being tested.

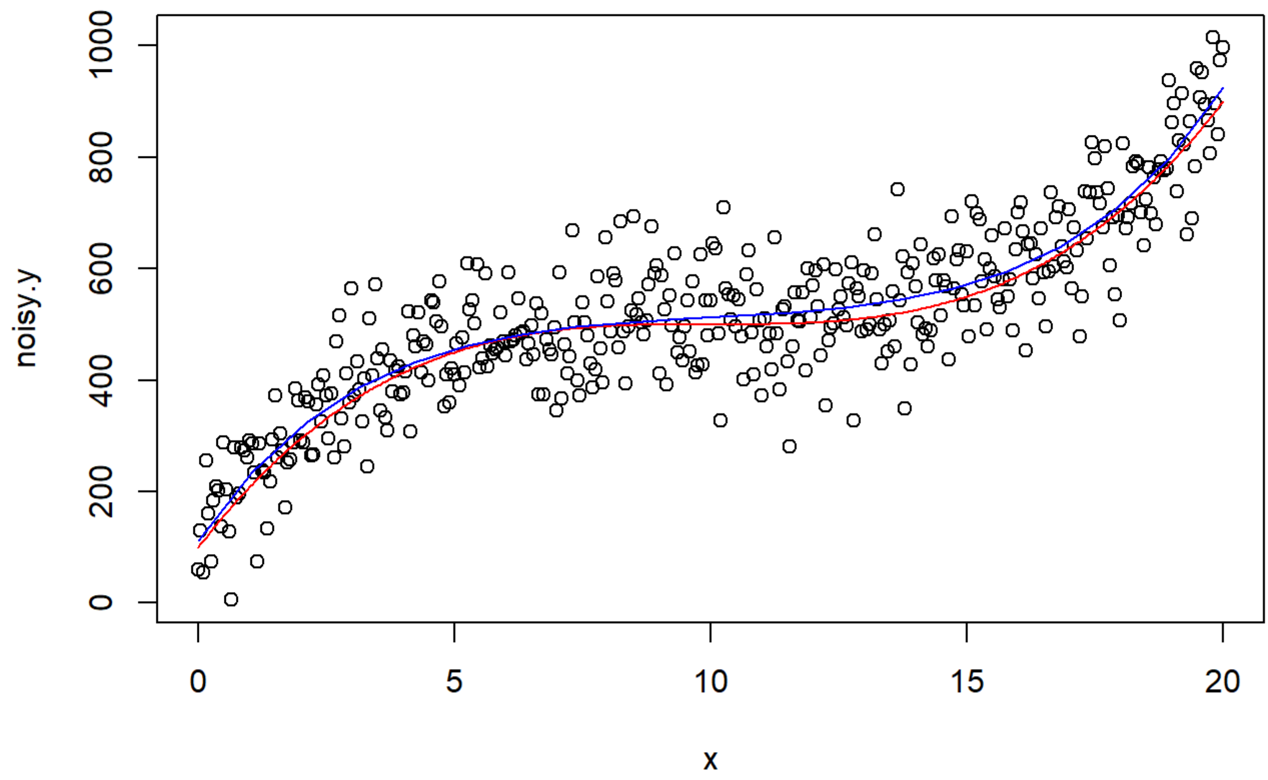
```
  summary_table <- data.frame(summary_degree=numeric(0), sum_error =numeric(0), count_error=n
umeric(0), mean_error=numeric(0))

  a <- 0
  b <- 0
  c <- 0

  for (k in 1:5){
    for (j in 1:nrow(results)){
      ifelse(results[j,1] == k, a <- a + results[j,3], a <- a)
      ifelse(results[j,1] == k, b <- b + 1, b <- b)
    }
    c = a / b
    row <- data.frame(summary_degree = k, sum_error = a, count_error = b, mean_error = c)
    summary_table <- rbind(summary_table, row)
    a = 0
    b = 0
    c = 0
  }
summary_table
```

```
##   summary_degree sum_error count_error mean_error
## 1              1  91824.79          10   9182.479
## 2              2  91824.62          10   9182.462
## 3              3  60496.24          10   6049.624
## 4              4  60495.08          10   6049.508
## 5              5  60216.85          10   6021.685
```

1c. Plot the best model's fitted line in blue and compare to the true function (the red line from the previous plot).

From the table above, the model with the lowest MSE (and therefore the best model) is the degree 5 polynomial.

```
plot(x,noisy.y)
lines(x, y, col='red')
poly5_mod <-lm(noisy.y ~ poly(x, 5, raw = T), data = df_train)
x <- seq(min(df$x), max(df$x), length.out=20)
y <- predict(poly5_mod, newdata = data.frame(x = x))
lines(x, y, col = "blue")
```



1.d Comment on the results of (c). Why was performance better or worse at different order polynomials?

Since we generated the data, we know that the true underlying function is a cubic function (with noise added). My 10-fold cross-validation shows that the degree 5 polynomial has the lowest cross-validation error (though this error value is very similar to the polynomials of degrees 3 and 4). We can see from this plot that the 5 degree polynomial model is still visually a good fit to the data points, and stays fairly close to the true function.

We can hazard that the reason a degree 5 model has slightly lower error than degree 3 model may be due to the noise we added to the true function, therefore the 4 and 5 degree models are over-fitting to the noise as the higher the degree of polynomial, the more flexible it is. Cross-validation is only an approximation of the test error so it is possible for over-fitting to occur, however this is more common when the sample size is small.

1e. Report the CV error and test error at each order of polynomial. Which achieves the lowest CV error? How does the

CV error compare to the test error? Comment on the results.

```
degree <- 1:5
test_results <- data.frame(degree=numeric(0), test_error=numeric(0))
for (d in degree){
    mod <- glm(noisy.y ~ poly(x, d, raw = T), data=df_test)
    test_row <- data.frame(degree=d,     test_error = mean((predict(mod, df_test) - df_test$n
oisy.y)^2))
    test_results <- rbind(test_results, test_row)
    }
test_results <- cbind(test_results, summary_table[,4, drop=FALSE])
names(test_results)[names(test_results) == "mean_error"] <- "cross-validation_error"
test_results
```

```
##   degree test_error cross-validation_error
## 1      1  10474.458               9182.479
## 2      2  10299.480               9182.462
## 3      3   5596.121               6049.624
## 4      4   5560.176               6049.508
## 5      5   5558.305               6021.685
```

For both test error and cross-validation error, the error decreases as the degree of the polynomial increases (from 1 through to 5).

The lowest cross-validation error and test error are both for the polynomial of degree 5.

While test error is higher than validation error for degrees 1 and 2, it is lower for degrees 3, 4 and 5. It may be that the test set contains more “easy” y’s to predict based on x’s than in the validation set, so the model performs better at degrees 3, 4 and 5 on test error, compared to cross-validation. However, since my randomisation of the dataset into a training and test split was intended to try to mitigate this issue.

2a. Pick a new dataset from the `mlbench` package (one we haven’t used in class that is 2-dimensional with two classes). Experiment with classifying the data using KNN at different values of k. Use cross-validation to choose your best model.

```
library(mlbench)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

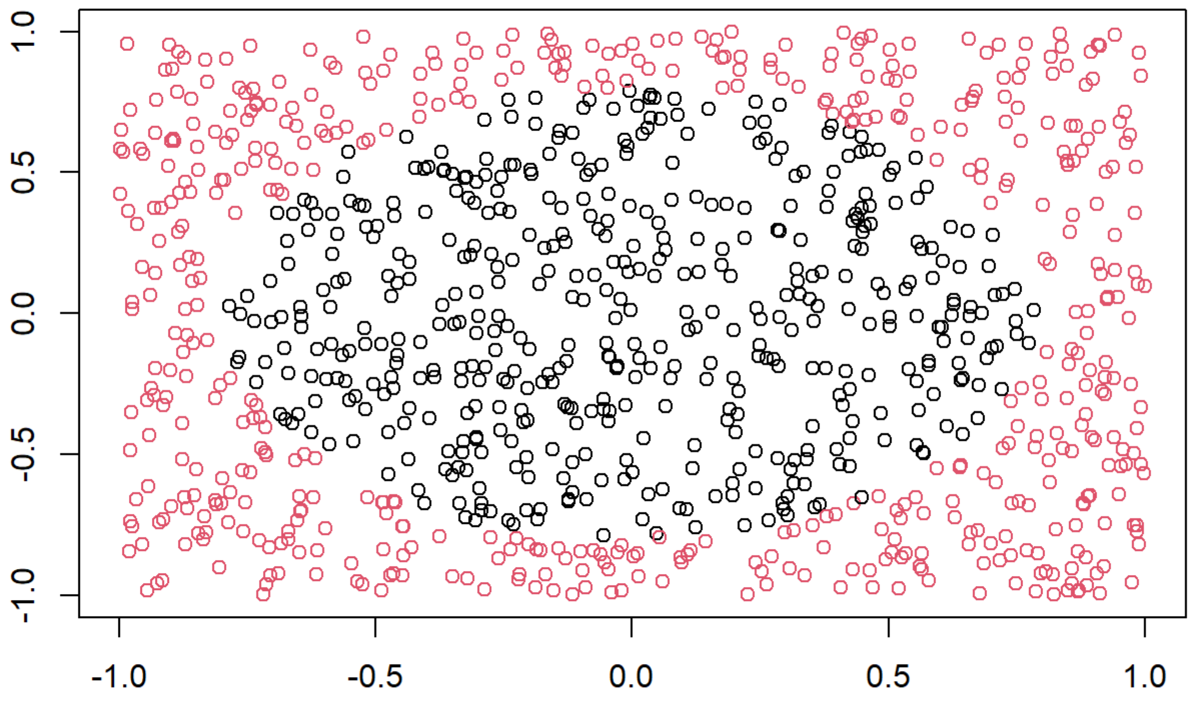
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
set.seed(1)
ls(package:mlbench)
```

```
##  [1] "bayesclass"        "mlbench.1spiral"   "mlbench.2dnormals"
##  [4] "mlbench.cassini"   "mlbench.circle"    "mlbench.corners"
##  [7] "mlbench.cuboids"   "mlbench.friedman1" "mlbench.friedman2"
## [10] "mlbench.friedman3" "mlbench.hypercube" "mlbench.peak"
## [13] "mlbench.ringnorm"  "mlbench.shapes"    "mlbench.simplex"
## [16] "mlbench.smiley"    "mlbench.spirals"   "mlbench.threenorm"
## [19] "mlbench.twonorm"   "mlbench.waveform"  "mlbench.xor"
```

```
c <- mlbench.circle(1000,2)
plot(c)
```



Move the dataset into a dataframe so it’s easier to work with. Rename variables and set levels of classes.

```
cdf <- data.frame(c)
cdf <- cdf %>% rename(Y.1 = classes, X.1 = x.1, X.2 = x.2)
levels(cdf$Y.1) <- c("0", "1")
```

Randomly permute the training set and split into 10 evenly sized folds. Then create a for loop: for values 1 through 100 of K, split the training set again into a training set and test set, with 1 fold in the test set and the rest in the training set. Change which fold is in the test set each time by inserting a nested for loop for i (fold used for test) 1 through 10 (for 10-fold cross-validation). Calculate the mean-squared error of the test sub-set (validation sub-set) for each fold. Store results from each loop in a results table with column headings for the value of the k parameter, the fold number and the classification error.

```
library(class)
n_test <- floor(nrow(cdf)*0.2)
idx <- sample(1:nrow(cdf), n_test)
train <- cdf[-idx,]
test <- cdf[idx,]
folds <- sample(rep(1:10, each = nrow(train)/10), replace=FALSE)
results <- data.frame(results_k=numeric(0), results_i=numeric(0), results_error=numeric(0))
range <- 1:100
for (k in range){
  for (i in 1:10){
  cv.te <- train[folds==i,]
  cv.tr <- train[folds!=i,]
  cv.tr.x <- cv.tr[,1:2]
  cv.tr.y <- cv.tr[,3]
  cv.te.x <- cv.te[,1:2]
  cv.te.y <- cv.te[,3]
  pred.Y <- knn(cv.tr.x, cv.te.x, cv.tr.y, k)
  row <- data.frame(results_k=k, results_i=i, results_error=mean(cv.te.y != pred.Y))
  results <- rbind(results, row)
  }
}
results
```

```
##      results_k results_i results_error
## 1           1         1        0.0625
## 2           1         2        0.0000
## 3           1         3        0.0375
## 4           1         4        0.0125
## 5           1         5        0.0250
## 6           1         6        0.0250
## 7           1         7        0.0000
## 8           1         8        0.0125
## 9           1         9        0.0250
## 10          1        10        0.0125
## 11          2         1        0.1000
## 12          2         2        0.0000
## 13          2         3        0.0500
## 14          2         4        0.0500
## 15          2         5        0.0375
## 16          2         6        0.0500
## 17          2         7        0.0000
## 18          2         8        0.0250
## 19          2         9        0.0250
## 20          2        10        0.0250
## 21          3         1        0.0625
## 22          3         2        0.0125
## 23          3         3        0.0250
## 24          3         4        0.0625
## 25          3         5        0.0250
## 26          3         6        0.0250
## 27          3         7        0.0125
## 28          3         8        0.0250
## 29          3         9        0.0125
## 30          3        10        0.0375
## 31          4         1        0.0625
## 32          4         2        0.0250
## 33          4         3        0.0375
## 34          4         4        0.0750
## 35          4         5        0.0000
## 36          4         6        0.0250
## 37          4         7        0.0000
## 38          4         8        0.0375
## 39          4         9        0.0375
## 40          4        10        0.0500
## 41          5         1        0.0500
## 42          5         2        0.0125
## 43          5         3        0.0125
## 44          5         4        0.0875
## 45          5         5        0.0250
## 46          5         6        0.0250
## 47          5         7        0.0125
## 48          5         8        0.0250
## 49          5         9        0.0125
## 50          5        10        0.0250
## 51          6         1        0.0625
## 52          6         2        0.0250
## 53          6         3        0.0125
## 54          6         4        0.0625
```

```
## 55        6     5     0.0250
## 56        6     6     0.0250
## 57        6     7     0.0250
## 58        6     8     0.0250
## 59        6     9     0.0125
## 60        6    10     0.0500
## 61        7     1     0.0375
## 62        7     2     0.0375
## 63        7     3     0.0250
## 64        7     4     0.0750
## 65        7     5     0.0500
## 66        7     6     0.0250
## 67        7     7     0.0125
## 68        7     8     0.0250
## 69        7     9     0.0375
## 70        7    10     0.0250
## 71        8     1     0.0500
## 72        8     2     0.0500
## 73        8     3     0.0250
## 74        8     4     0.0750
## 75        8     5     0.0500
## 76        8     6     0.0375
## 77        8     7     0.0250
## 78        8     8     0.0125
## 79        8     9     0.0500
## 80        8    10     0.0375
## 81        9     1     0.0375
## 82        9     2     0.0625
## 83        9     3     0.0125
## 84        9     4     0.0875
## 85        9     5     0.0500
## 86        9     6     0.0375
## 87        9     7     0.0250
## 88        9     8     0.0250
## 89        9     9     0.0500
## 90        9    10     0.0375
## 91       10     1     0.0375
## 92       10     2     0.0625
## 93       10     3     0.0250
## 94       10     4     0.0750
## 95       10     5     0.0375
## 96       10     6     0.0750
## 97       10     7     0.0250
## 98       10     8     0.0250
## 99       10     9     0.0500
## 100      10    10     0.0250
## 101      11     1     0.0250
## 102      11     2     0.0500
## 103      11     3     0.0250
## 104      11     4     0.0625
## 105      11     5     0.0250
## 106      11     6     0.0500
## 107      11     7     0.0250
## 108      11     8     0.0250
## 109      11     9     0.0500
## 110      11    10     0.0250
```

```
## 111          12          1        0.0375
## 112          12          2        0.0500
## 113          12          3        0.0375
## 114          12          4        0.0500
## 115          12          5        0.0500
## 116          12          6        0.0500
## 117          12          7        0.0250
## 118          12          8        0.0375
## 119          12          9        0.0625
## 120          12         10        0.0250
## 121          13          1        0.0250
## 122          13          2        0.0500
## 123          13          3        0.0250
## 124          13          4        0.0500
## 125          13          5        0.0250
## 126          13          6        0.0375
## 127          13          7        0.0250
## 128          13          8        0.0375
## 129          13          9        0.0375
## 130          13         10        0.0250
## 131          14          1        0.0500
## 132          14          2        0.0250
## 133          14          3        0.0375
## 134          14          4        0.0500
## 135          14          5        0.0250
## 136          14          6        0.0375
## 137          14          7        0.0250
## 138          14          8        0.0500
## 139          14          9        0.0375
## 140          14         10        0.0375
## 141          15          1        0.0250
## 142          15          2        0.0250
## 143          15          3        0.0250
## 144          15          4        0.0375
## 145          15          5        0.0125
## 146          15          6        0.0375
## 147          15          7        0.0250
## 148          15          8        0.0375
## 149          15          9        0.0250
## 150          15         10        0.0250
## 151          16          1        0.0250
## 152          16          2        0.0250
## 153          16          3        0.0375
## 154          16          4        0.0625
## 155          16          5        0.0125
## 156          16          6        0.0375
## 157          16          7        0.0250
## 158          16          8        0.0125
## 159          16          9        0.0250
## 160          16         10        0.0125
## 161          17          1        0.0375
## 162          17          2        0.0500
## 163          17          3        0.0250
## 164          17          4        0.0500
## 165          17          5        0.0125
## 166          17          6        0.0250
```

```
## 167        17        7        0.0250
## 168        17        8        0.0250
## 169        17        9        0.0250
## 170        17       10        0.0250
## 171        18        1        0.0250
## 172        18        2        0.0500
## 173        18        3        0.0250
## 174        18        4        0.0375
## 175        18        5        0.0250
## 176        18        6        0.0250
## 177        18        7        0.0125
## 178        18        8        0.0375
## 179        18        9        0.0250
## 180        18       10        0.0250
## 181        19        1        0.0375
## 182        19        2        0.0375
## 183        19        3        0.0250
## 184        19        4        0.0375
## 185        19        5        0.0125
## 186        19        6        0.0250
## 187        19        7        0.0125
## 188        19        8        0.0500
## 189        19        9        0.0375
## 190        19       10        0.0375
## 191        20        1        0.0375
## 192        20        2        0.0250
## 193        20        3        0.0125
## 194        20        4        0.0250
## 195        20        5        0.0500
## 196        20        6        0.0250
## 197        20        7        0.0125
## 198        20        8        0.0250
## 199        20        9        0.0250
## 200        20       10        0.0500
## 201        21        1        0.0250
## 202        21        2        0.0375
## 203        21        3        0.0250
## 204        21        4        0.0375
## 205        21        5        0.0250
## 206        21        6        0.0250
## 207        21        7        0.0250
## 208        21        8        0.0375
## 209        21        9        0.0375
## 210        21       10        0.0500
## 211        22        1        0.0375
## 212        22        2        0.0500
## 213        22        3        0.0250
## 214        22        4        0.0500
## 215        22        5        0.0000
## 216        22        6        0.0625
## 217        22        7        0.0375
## 218        22        8        0.0375
## 219        22        9        0.0250
## 220        22       10        0.0500
## 221        23        1        0.0250
## 222        23        2        0.0375
```

```
## 223          23          3          0.0250
## 224          23          4          0.0500
## 225          23          5          0.0125
## 226          23          6          0.0375
## 227          23          7          0.0250
## 228          23          8          0.0375
## 229          23          9          0.0250
## 230          23         10          0.0500
## 231          24          1          0.0375
## 232          24          2          0.0375
## 233          24          3          0.0250
## 234          24          4          0.0500
## 235          24          5          0.0375
## 236          24          6          0.0375
## 237          24          7          0.0125
## 238          24          8          0.0500
## 239          24          9          0.0375
## 240          24         10          0.0375
## 241          25          1          0.0250
## 242          25          2          0.0375
## 243          25          3          0.0375
## 244          25          4          0.0375
## 245          25          5          0.0250
## 246          25          6          0.0375
## 247          25          7          0.0125
## 248          25          8          0.0625
## 249          25          9          0.0375
## 250          25         10          0.0500
## 251          26          1          0.0250
## 252          26          2          0.0625
## 253          26          3          0.0375
## 254          26          4          0.0375
## 255          26          5          0.0500
## 256          26          6          0.0500
## 257          26          7          0.0125
## 258          26          8          0.0375
## 259          26          9          0.0250
## 260          26         10          0.0250
## 261          27          1          0.0375
## 262          27          2          0.0500
## 263          27          3          0.0375
## 264          27          4          0.0375
## 265          27          5          0.0250
## 266          27          6          0.0375
## 267          27          7          0.0000
## 268          27          8          0.0750
## 269          27          9          0.0125
## 270          27         10          0.0500
## 271          28          1          0.0625
## 272          28          2          0.0500
## 273          28          3          0.0375
## 274          28          4          0.0500
## 275          28          5          0.0375
## 276          28          6          0.0375
## 277          28          7          0.0125
## 278          28          8          0.0750
```

```
## 279          28         9          0.0125
## 280          28        10          0.0625
## 281          29         1          0.0375
## 282          29         2          0.0500
## 283          29         3          0.0375
## 284          29         4          0.0375
## 285          29         5          0.0250
## 286          29         6          0.0500
## 287          29         7          0.0000
## 288          29         8          0.0750
## 289          29         9          0.0125
## 290          29        10          0.0375
## 291          30         1          0.0625
## 292          30         2          0.0625
## 293          30         3          0.0375
## 294          30         4          0.0375
## 295          30         5          0.0375
## 296          30         6          0.0500
## 297          30         7          0.0125
## 298          30         8          0.0875
## 299          30         9          0.0125
## 300          30        10          0.0375
## 301          31         1          0.0375
## 302          31         2          0.0625
## 303          31         3          0.0375
## 304          31         4          0.0500
## 305          31         5          0.0375
## 306          31         6          0.0500
## 307          31         7          0.0125
## 308          31         8          0.0750
## 309          31         9          0.0125
## 310          31        10          0.0375
## 311          32         1          0.0500
## 312          32         2          0.0625
## 313          32         3          0.0500
## 314          32         4          0.0375
## 315          32         5          0.0375
## 316          32         6          0.0500
## 317          32         7          0.0000
## 318          32         8          0.0875
## 319          32         9          0.0125
## 320          32        10          0.0500
## 321          33         1          0.0625
## 322          33         2          0.0625
## 323          33         3          0.0375
## 324          33         4          0.0375
## 325          33         5          0.0375
## 326          33         6          0.0500
## 327          33         7          0.0000
## 328          33         8          0.0750
## 329          33         9          0.0125
## 330          33        10          0.0250
## 331          34         1          0.0625
## 332          34         2          0.0625
## 333          34         3          0.0375
## 334          34         4          0.0250
```

```
## 335          34       5         0.0375
## 336          34       6         0.0500
## 337          34       7         0.0250
## 338          34       8         0.0750
## 339          34       9         0.0250
## 340          34      10         0.0375
## 341          35       1         0.0500
## 342          35       2         0.0750
## 343          35       3         0.0500
## 344          35       4         0.0375
## 345          35       5         0.0500
## 346          35       6         0.0500
## 347          35       7         0.0000
## 348          35       8         0.0625
## 349          35       9         0.0250
## 350          35      10         0.0375
## 351          36       1         0.0750
## 352          36       2         0.0750
## 353          36       3         0.0625
## 354          36       4         0.0500
## 355          36       5         0.0500
## 356          36       6         0.0500
## 357          36       7         0.0125
## 358          36       8         0.0500
## 359          36       9         0.0125
## 360          36      10         0.0375
## 361          37       1         0.0875
## 362          37       2         0.0750
## 363          37       3         0.0625
## 364          37       4         0.0500
## 365          37       5         0.0500
## 366          37       6         0.0500
## 367          37       7         0.0000
## 368          37       8         0.0625
## 369          37       9         0.0125
## 370          37      10         0.0500
## 371          38       1         0.0500
## 372          38       2         0.0750
## 373          38       3         0.0625
## 374          38       4         0.0375
## 375          38       5         0.0500
## 376          38       6         0.0500
## 377          38       7         0.0000
## 378          38       8         0.0625
## 379          38       9         0.0250
## 380          38      10         0.0500
## 381          39       1         0.0500
## 382          39       2         0.0875
## 383          39       3         0.0500
## 384          39       4         0.0500
## 385          39       5         0.0500
## 386          39       6         0.0500
## 387          39       7         0.0000
## 388          39       8         0.0625
## 389          39       9         0.0125
## 390          39      10         0.0375
```

```
## 391          40         1         0.0125
## 392          40         2         0.0625
## 393          40         3         0.0500
## 394          40         4         0.0500
## 395          40         5         0.0500
## 396          40         6         0.0500
## 397          40         7         0.0000
## 398          40         8         0.0625
## 399          40         9         0.0375
## 400          40        10         0.0500
## 401          41         1         0.0375
## 402          41         2         0.0625
## 403          41         3         0.0625
## 404          41         4         0.0250
## 405          41         5         0.0500
## 406          41         6         0.0625
## 407          41         7         0.0000
## 408          41         8         0.0625
## 409          41         9         0.0125
## 410          41        10         0.0375
## 411          42         1         0.0500
## 412          42         2         0.0875
## 413          42         3         0.0625
## 414          42         4         0.0125
## 415          42         5         0.0500
## 416          42         6         0.0625
## 417          42         7         0.0000
## 418          42         8         0.0750
## 419          42         9         0.0250
## 420          42        10         0.0500
## 421          43         1         0.0500
## 422          43         2         0.0750
## 423          43         3         0.0875
## 424          43         4         0.0250
## 425          43         5         0.0375
## 426          43         6         0.0625
## 427          43         7         0.0000
## 428          43         8         0.0625
## 429          43         9         0.0125
## 430          43        10         0.0500
## 431          44         1         0.0500
## 432          44         2         0.0750
## 433          44         3         0.0875
## 434          44         4         0.0375
## 435          44         5         0.0250
## 436          44         6         0.0625
## 437          44         7         0.0000
## 438          44         8         0.0625
## 439          44         9         0.0250
## 440          44        10         0.0625
## 441          45         1         0.0375
## 442          45         2         0.0750
## 443          45         3         0.0875
## 444          45         4         0.0250
## 445          45         5         0.0375
## 446          45         6         0.0625
```

```
## 447          45       7          0.0000
## 448          45       8          0.0750
## 449          45       9          0.0125
## 450          45      10          0.0625
## 451          46       1          0.0625
## 452          46       2          0.0875
## 453          46       3          0.0875
## 454          46       4          0.0250
## 455          46       5          0.0375
## 456          46       6          0.0625
## 457          46       7          0.0125
## 458          46       8          0.0750
## 459          46       9          0.0250
## 460          46      10          0.0750
## 461          47       1          0.0500
## 462          47       2          0.0875
## 463          47       3          0.0875
## 464          47       4          0.0250
## 465          47       5          0.0500
## 466          47       6          0.0625
## 467          47       7          0.0125
## 468          47       8          0.0750
## 469          47       9          0.0125
## 470          47      10          0.0625
## 471          48       1          0.0375
## 472          48       2          0.0875
## 473          48       3          0.0875
## 474          48       4          0.0250
## 475          48       5          0.0250
## 476          48       6          0.0625
## 477          48       7          0.0125
## 478          48       8          0.0875
## 479          48       9          0.0250
## 480          48      10          0.0750
## 481          49       1          0.0375
## 482          49       2          0.0875
## 483          49       3          0.0875
## 484          49       4          0.0375
## 485          49       5          0.0500
## 486          49       6          0.0625
## 487          49       7          0.0125
## 488          49       8          0.0875
## 489          49       9          0.0125
## 490          49      10          0.1000
## 491          50       1          0.0625
## 492          50       2          0.0750
## 493          50       3          0.0875
## 494          50       4          0.0375
## 495          50       5          0.0500
## 496          50       6          0.0625
## 497          50       7          0.0125
## 498          50       8          0.0875
## 499          50       9          0.0125
## 500          50      10          0.0875
## 501          51       1          0.0625
## 502          51       2          0.0875
```

```
## 503          51        3         0.0750
## 504          51        4         0.0250
## 505          51        5         0.0375
## 506          51        6         0.0625
## 507          51        7         0.0125
## 508          51        8         0.0875
## 509          51        9         0.0125
## 510          51       10         0.0750
## 511          52        1         0.0625
## 512          52        2         0.0875
## 513          52        3         0.0625
## 514          52        4         0.0500
## 515          52        5         0.0500
## 516          52        6         0.0500
## 517          52        7         0.0250
## 518          52        8         0.0875
## 519          52        9         0.0125
## 520          52       10         0.0750
## 521          53        1         0.0625
## 522          53        2         0.0875
## 523          53        3         0.0750
## 524          53        4         0.0250
## 525          53        5         0.0375
## 526          53        6         0.0500
## 527          53        7         0.0375
## 528          53        8         0.0875
## 529          53        9         0.0125
## 530          53       10         0.0750
## 531          54        1         0.0625
## 532          54        2         0.0875
## 533          54        3         0.0625
## 534          54        4         0.0375
## 535          54        5         0.0375
## 536          54        6         0.0500
## 537          54        7         0.0375
## 538          54        8         0.0875
## 539          54        9         0.0125
## 540          54       10         0.0750
## 541          55        1         0.0625
## 542          55        2         0.0875
## 543          55        3         0.1000
## 544          55        4         0.0375
## 545          55        5         0.0625
## 546          55        6         0.0750
## 547          55        7         0.0375
## 548          55        8         0.0750
## 549          55        9         0.0000
## 550          55       10         0.0750
## 551          56        1         0.0750
## 552          56        2         0.0875
## 553          56        3         0.1125
## 554          56        4         0.0375
## 555          56        5         0.0500
## 556          56        6         0.0750
## 557          56        7         0.0375
## 558          56        8         0.0750
```

```
## 559          56        9          0.0000
## 560          56       10          0.0750
## 561          57        1          0.0750
## 562          57        2          0.0875
## 563          57        3          0.0875
## 564          57        4          0.0250
## 565          57        5          0.0500
## 566          57        6          0.0875
## 567          57        7          0.0375
## 568          57        8          0.0875
## 569          57        9          0.0000
## 570          57       10          0.0750
## 571          58        1          0.0875
## 572          58        2          0.0875
## 573          58        3          0.0875
## 574          58        4          0.0375
## 575          58        5          0.0500
## 576          58        6          0.0750
## 577          58        7          0.0375
## 578          58        8          0.0875
## 579          58        9          0.0000
## 580          58       10          0.0750
## 581          59        1          0.0750
## 582          59        2          0.0750
## 583          59        3          0.1000
## 584          59        4          0.0375
## 585          59        5          0.0625
## 586          59        6          0.0750
## 587          59        7          0.0375
## 588          59        8          0.0875
## 589          59        9          0.0125
## 590          59       10          0.0750
## 591          60        1          0.0625
## 592          60        2          0.0875
## 593          60        3          0.0875
## 594          60        4          0.0375
## 595          60        5          0.0625
## 596          60        6          0.0875
## 597          60        7          0.0375
## 598          60        8          0.0875
## 599          60        9          0.0125
## 600          60       10          0.0750
## 601          61        1          0.1000
## 602          61        2          0.0875
## 603          61        3          0.0875
## 604          61        4          0.0375
## 605          61        5          0.0625
## 606          61        6          0.1000
## 607          61        7          0.0250
## 608          61        8          0.0875
## 609          61        9          0.0125
## 610          61       10          0.0750
## 611          62        1          0.0875
## 612          62        2          0.0750
## 613          62        3          0.1000
## 614          62        4          0.0500
```

```
## 615          62        5         0.0625
## 616          62        6         0.0750
## 617          62        7         0.0375
## 618          62        8         0.0750
## 619          62        9         0.0125
## 620          62       10         0.1000
## 621          63        1         0.0875
## 622          63        2         0.0750
## 623          63        3         0.1125
## 624          63        4         0.0250
## 625          63        5         0.0625
## 626          63        6         0.1000
## 627          63        7         0.0375
## 628          63        8         0.0875
## 629          63        9         0.0125
## 630          63       10         0.1000
## 631          64        1         0.0875
## 632          64        2         0.1000
## 633          64        3         0.1125
## 634          64        4         0.0500
## 635          64        5         0.0625
## 636          64        6         0.1000
## 637          64        7         0.0375
## 638          64        8         0.0875
## 639          64        9         0.0125
## 640          64       10         0.1000
## 641          65        1         0.0875
## 642          65        2         0.0875
## 643          65        3         0.1000
## 644          65        4         0.0500
## 645          65        5         0.0625
## 646          65        6         0.0875
## 647          65        7         0.0500
## 648          65        8         0.0875
## 649          65        9         0.0125
## 650          65       10         0.1125
## 651          66        1         0.0875
## 652          66        2         0.1125
## 653          66        3         0.1125
## 654          66        4         0.0250
## 655          66        5         0.0625
## 656          66        6         0.0875
## 657          66        7         0.0500
## 658          66        8         0.0750
## 659          66        9         0.0125
## 660          66       10         0.1125
## 661          67        1         0.0875
## 662          67        2         0.1125
## 663          67        3         0.1000
## 664          67        4         0.0375
## 665          67        5         0.0625
## 666          67        6         0.1000
## 667          67        7         0.0375
## 668          67        8         0.0875
## 669          67        9         0.0125
## 670          67       10         0.1125
```

```
## 671          68         1          0.0750
## 672          68         2          0.1125
## 673          68         3          0.0875
## 674          68         4          0.0375
## 675          68         5          0.0625
## 676          68         6          0.1000
## 677          68         7          0.0500
## 678          68         8          0.1000
## 679          68         9          0.0125
## 680          68        10          0.1250
## 681          69         1          0.0875
## 682          69         2          0.1000
## 683          69         3          0.0875
## 684          69         4          0.0375
## 685          69         5          0.0625
## 686          69         6          0.1000
## 687          69         7          0.0375
## 688          69         8          0.0875
## 689          69         9          0.0125
## 690          69        10          0.1250
## 691          70         1          0.0875
## 692          70         2          0.0875
## 693          70         3          0.1000
## 694          70         4          0.0375
## 695          70         5          0.0750
## 696          70         6          0.1000
## 697          70         7          0.0375
## 698          70         8          0.0875
## 699          70         9          0.0125
## 700          70        10          0.1250
## 701          71         1          0.0875
## 702          71         2          0.1125
## 703          71         3          0.0875
## 704          71         4          0.0375
## 705          71         5          0.0750
## 706          71         6          0.1000
## 707          71         7          0.0375
## 708          71         8          0.0750
## 709          71         9          0.0250
## 710          71        10          0.1250
## 711          72         1          0.0875
## 712          72         2          0.1125
## 713          72         3          0.0875
## 714          72         4          0.0500
## 715          72         5          0.0750
## 716          72         6          0.0875
## 717          72         7          0.0375
## 718          72         8          0.0750
## 719          72         9          0.0250
## 720          72        10          0.1125
## 721          73         1          0.1000
## 722          73         2          0.1250
## 723          73         3          0.0875
## 724          73         4          0.0500
## 725          73         5          0.0750
## 726          73         6          0.1125
```

```
## 727          73         7         0.0375
## 728          73         8         0.0875
## 729          73         9         0.0375
## 730          73        10         0.1375
## 731          74         1         0.1125
## 732          74         2         0.1125
## 733          74         3         0.1000
## 734          74         4         0.0500
## 735          74         5         0.0750
## 736          74         6         0.1000
## 737          74         7         0.0375
## 738          74         8         0.0875
## 739          74         9         0.0375
## 740          74        10         0.1125
## 741          75         1         0.1125
## 742          75         2         0.1125
## 743          75         3         0.1000
## 744          75         4         0.0500
## 745          75         5         0.0750
## 746          75         6         0.1000
## 747          75         7         0.0375
## 748          75         8         0.0875
## 749          75         9         0.0375
## 750          75        10         0.1250
## 751          76         1         0.1000
## 752          76         2         0.1250
## 753          76         3         0.1000
## 754          76         4         0.0500
## 755          76         5         0.0750
## 756          76         6         0.0875
## 757          76         7         0.0500
## 758          76         8         0.1000
## 759          76         9         0.0375
## 760          76        10         0.1125
## 761          77         1         0.1125
## 762          77         2         0.1125
## 763          77         3         0.1000
## 764          77         4         0.0500
## 765          77         5         0.0750
## 766          77         6         0.0875
## 767          77         7         0.0375
## 768          77         8         0.1000
## 769          77         9         0.0250
## 770          77        10         0.1000
## 771          78         1         0.1125
## 772          78         2         0.1250
## 773          78         3         0.1000
## 774          78         4         0.0500
## 775          78         5         0.0750
## 776          78         6         0.0875
## 777          78         7         0.0500
## 778          78         8         0.1000
## 779          78         9         0.0375
## 780          78        10         0.1125
## 781          79         1         0.1125
## 782          79         2         0.1250
```

```
## 783         79          3          0.1000
## 784         79          4          0.0500
## 785         79          5          0.0750
## 786         79          6          0.0875
## 787         79          7          0.0375
## 788         79          8          0.1125
## 789         79          9          0.0500
## 790         79         10          0.1250
## 791         80          1          0.1125
## 792         80          2          0.1250
## 793         80          3          0.1000
## 794         80          4          0.0375
## 795         80          5          0.0750
## 796         80          6          0.0875
## 797         80          7          0.0500
## 798         80          8          0.1125
## 799         80          9          0.0375
## 800         80         10          0.1000
## 801         81          1          0.1125
## 802         81          2          0.1250
## 803         81          3          0.0875
## 804         81          4          0.0500
## 805         81          5          0.0750
## 806         81          6          0.0875
## 807         81          7          0.0375
## 808         81          8          0.1125
## 809         81          9          0.0500
## 810         81         10          0.1375
## 811         82          1          0.1125
## 812         82          2          0.1250
## 813         82          3          0.1000
## 814         82          4          0.0500
## 815         82          5          0.0750
## 816         82          6          0.0875
## 817         82          7          0.0375
## 818         82          8          0.1125
## 819         82          9          0.0500
## 820         82         10          0.1375
## 821         83          1          0.1000
## 822         83          2          0.1500
## 823         83          3          0.1000
## 824         83          4          0.0625
## 825         83          5          0.0875
## 826         83          6          0.0875
## 827         83          7          0.0375
## 828         83          8          0.1125
## 829         83          9          0.0500
## 830         83         10          0.1125
## 831         84          1          0.1125
## 832         84          2          0.1375
## 833         84          3          0.1125
## 834         84          4          0.0750
## 835         84          5          0.0875
## 836         84          6          0.0875
## 837         84          7          0.0500
## 838         84          8          0.1125
```

```
## 839          84       9          0.0500
## 840          84      10          0.1500
## 841          85       1          0.1000
## 842          85       2          0.1375
## 843          85       3          0.1125
## 844          85       4          0.0625
## 845          85       5          0.0875
## 846          85       6          0.0750
## 847          85       7          0.0500
## 848          85       8          0.1250
## 849          85       9          0.0500
## 850          85      10          0.1250
## 851          86       1          0.1000
## 852          86       2          0.1375
## 853          86       3          0.1125
## 854          86       4          0.0500
## 855          86       5          0.0875
## 856          86       6          0.0875
## 857          86       7          0.0500
## 858          86       8          0.1125
## 859          86       9          0.0500
## 860          86      10          0.1250
## 861          87       1          0.1000
## 862          87       2          0.1375
## 863          87       3          0.1125
## 864          87       4          0.0750
## 865          87       5          0.0875
## 866          87       6          0.0750
## 867          87       7          0.0250
## 868          87       8          0.1125
## 869          87       9          0.0500
## 870          87      10          0.1125
## 871          88       1          0.1000
## 872          88       2          0.1375
## 873          88       3          0.1125
## 874          88       4          0.0750
## 875          88       5          0.0875
## 876          88       6          0.0750
## 877          88       7          0.0500
## 878          88       8          0.1250
## 879          88       9          0.0500
## 880          88      10          0.1250
## 881          89       1          0.1000
## 882          89       2          0.1375
## 883          89       3          0.1125
## 884          89       4          0.0625
## 885          89       5          0.1000
## 886          89       6          0.0625
## 887          89       7          0.0625
## 888          89       8          0.1250
## 889          89       9          0.0500
## 890          89      10          0.1250
## 891          90       1          0.1000
## 892          90       2          0.1375
## 893          90       3          0.1250
## 894          90       4          0.0500
```

```
## 895          90         5          0.1000
## 896          90         6          0.0625
## 897          90         7          0.0625
## 898          90         8          0.1125
## 899          90         9          0.0500
## 900          90        10          0.1250
## 901          91         1          0.1000
## 902          91         2          0.1375
## 903          91         3          0.1250
## 904          91         4          0.0500
## 905          91         5          0.0875
## 906          91         6          0.0625
## 907          91         7          0.0625
## 908          91         8          0.1125
## 909          91         9          0.0500
## 910          91        10          0.1250
## 911          92         1          0.1125
## 912          92         2          0.1375
## 913          92         3          0.1375
## 914          92         4          0.0375
## 915          92         5          0.0875
## 916          92         6          0.0625
## 917          92         7          0.0625
## 918          92         8          0.1125
## 919          92         9          0.0625
## 920          92        10          0.1625
## 921          93         1          0.1250
## 922          93         2          0.1375
## 923          93         3          0.1250
## 924          93         4          0.0500
## 925          93         5          0.0875
## 926          93         6          0.0500
## 927          93         7          0.0625
## 928          93         8          0.1125
## 929          93         9          0.0375
## 930          93        10          0.1250
## 931          94         1          0.1250
## 932          94         2          0.1250
## 933          94         3          0.1125
## 934          94         4          0.0750
## 935          94         5          0.0875
## 936          94         6          0.0625
## 937          94         7          0.0625
## 938          94         8          0.1125
## 939          94         9          0.0375
## 940          94        10          0.1375
## 941          95         1          0.1125
## 942          95         2          0.1250
## 943          95         3          0.1250
## 944          95         4          0.0625
## 945          95         5          0.0875
## 946          95         6          0.0500
## 947          95         7          0.0625
## 948          95         8          0.1000
## 949          95         9          0.0375
## 950          95        10          0.1375
```

```
## 951          96         1          0.1125
## 952          96         2          0.1375
## 953          96         3          0.1250
## 954          96         4          0.0625
## 955          96         5          0.0875
## 956          96         6          0.0500
## 957          96         7          0.0625
## 958          96         8          0.1000
## 959          96         9          0.0375
## 960          96        10          0.1375
## 961          97         1          0.1000
## 962          97         2          0.1375
## 963          97         3          0.1375
## 964          97         4          0.0625
## 965          97         5          0.0875
## 966          97         6          0.0375
## 967          97         7          0.0625
## 968          97         8          0.0875
## 969          97         9          0.0375
## 970          97        10          0.1500
## 971          98         1          0.1000
## 972          98         2          0.1375
## 973          98         3          0.1500
## 974          98         4          0.0625
## 975          98         5          0.0875
## 976          98         6          0.0375
## 977          98         7          0.0875
## 978          98         8          0.0750
## 979          98         9          0.0375
## 980          98        10          0.1500
## 981          99         1          0.0875
## 982          99         2          0.1375
## 983          99         3          0.1375
## 984          99         4          0.0625
## 985          99         5          0.1000
## 986          99         6          0.0375
## 987          99         7          0.0750
## 988          99         8          0.0875
## 989          99         9          0.0375
## 990          99        10          0.1375
## 991         100         1          0.1000
## 992         100         2          0.1250
## 993         100         3          0.1375
## 994         100         4          0.0625
## 995         100         5          0.0875
## 996         100         6          0.0375
## 997         100         7          0.0750
## 998         100         8          0.0875
## 999         100         9          0.0500
## 1000        100        10          0.1375
```

```
  summary_table_k <- data.frame(sum_k=numeric(0), sum_error =numeric(0), count_error=numeric
(0), mean_error=numeric(0))

  a <- 0
  b <- 0
  c <- 0

  for (k in range){
    for (j in 1:nrow(results)){
      ifelse(results[j,1] == k, a <- a + results[j,3], a <- a)
      ifelse(results[j,1] == k, b <- b + 1, b <- b)
    }
    c = a / b
    row <- data.frame(sum_k = k, sum_error = a, count_error = b, mean_error = c)
    summary_table_k <- rbind(summary_table_k, row)
    a = 0
    b = 0
    c = 0
  }
summary_table_k
```

```
##    sum_k sum_error count_error mean_error
## 1      1    0.2125          10    0.02125
## 2      2    0.3625          10    0.03625
## 3      3    0.3000          10    0.03000
## 4      4    0.3500          10    0.03500
## 5      5    0.2875          10    0.02875
## 6      6    0.3250          10    0.03250
## 7      7    0.3500          10    0.03500
## 8      8    0.4125          10    0.04125
## 9      9    0.4250          10    0.04250
## 10    10    0.4375          10    0.04375
## 11    11    0.3625          10    0.03625
## 12    12    0.4250          10    0.04250
## 13    13    0.3375          10    0.03375
## 14    14    0.3750          10    0.03750
## 15    15    0.2750          10    0.02750
## 16    16    0.2750          10    0.02750
## 17    17    0.3000          10    0.03000
## 18    18    0.2875          10    0.02875
## 19    19    0.3125          10    0.03125
## 20    20    0.2875          10    0.02875
## 21    21    0.3250          10    0.03250
## 22    22    0.3750          10    0.03750
## 23    23    0.3250          10    0.03250
## 24    24    0.3625          10    0.03625
## 25    25    0.3625          10    0.03625
## 26    26    0.3625          10    0.03625
## 27    27    0.3625          10    0.03625
## 28    28    0.4375          10    0.04375
## 29    29    0.3625          10    0.03625
## 30    30    0.4375          10    0.04375
## 31    31    0.4125          10    0.04125
## 32    32    0.4375          10    0.04375
## 33    33    0.4000          10    0.04000
## 34    34    0.4375          10    0.04375
## 35    35    0.4375          10    0.04375
## 36    36    0.4750          10    0.04750
## 37    37    0.5000          10    0.05000
## 38    38    0.4625          10    0.04625
## 39    39    0.4500          10    0.04500
## 40    40    0.4250          10    0.04250
## 41    41    0.4125          10    0.04125
## 42    42    0.4750          10    0.04750
## 43    43    0.4625          10    0.04625
## 44    44    0.4875          10    0.04875
## 45    45    0.4750          10    0.04750
## 46    46    0.5500          10    0.05500
## 47    47    0.5250          10    0.05250
## 48    48    0.5250          10    0.05250
## 49    49    0.5750          10    0.05750
## 50    50    0.5750          10    0.05750
## 51    51    0.5375          10    0.05375
## 52    52    0.5625          10    0.05625
## 53    53    0.5500          10    0.05500
## 54    54    0.5500          10    0.05500
```

```
## 55     55    0.6125          10    0.06125
## 56     56    0.6250          10    0.06250
## 57     57    0.6125          10    0.06125
## 58     58    0.6250          10    0.06250
## 59     59    0.6375          10    0.06375
## 60     60    0.6375          10    0.06375
## 61     61    0.6750          10    0.06750
## 62     62    0.6750          10    0.06750
## 63     63    0.7000          10    0.07000
## 64     64    0.7500          10    0.07500
## 65     65    0.7375          10    0.07375
## 66     66    0.7375          10    0.07375
## 67     67    0.7500          10    0.07500
## 68     68    0.7625          10    0.07625
## 69     69    0.7375          10    0.07375
## 70     70    0.7500          10    0.07500
## 71     71    0.7625          10    0.07625
## 72     72    0.7500          10    0.07500
## 73     73    0.8500          10    0.08500
## 74     74    0.8250          10    0.08250
## 75     75    0.8375          10    0.08375
## 76     76    0.8375          10    0.08375
## 77     77    0.8000          10    0.08000
## 78     78    0.8500          10    0.08500
## 79     79    0.8750          10    0.08750
## 80     80    0.8375          10    0.08375
## 81     81    0.8750          10    0.08750
## 82     82    0.8875          10    0.08875
## 83     83    0.9000          10    0.09000
## 84     84    0.9750          10    0.09750
## 85     85    0.9250          10    0.09250
## 86     86    0.9125          10    0.09125
## 87     87    0.8875          10    0.08875
## 88     88    0.9375          10    0.09375
## 89     89    0.9375          10    0.09375
## 90     90    0.9250          10    0.09250
## 91     91    0.9125          10    0.09125
## 92     92    0.9750          10    0.09750
## 93     93    0.9125          10    0.09125
## 94     94    0.9375          10    0.09375
## 95     95    0.9000          10    0.09000
## 96     96    0.9125          10    0.09125
## 97     97    0.9000          10    0.09000
## 98     98    0.9250          10    0.09250
## 99     99    0.9000          10    0.09000
## 100   100    0.9000          10    0.09000
```

remove k = 1 as this is unhelpful. Then find the minimum error.

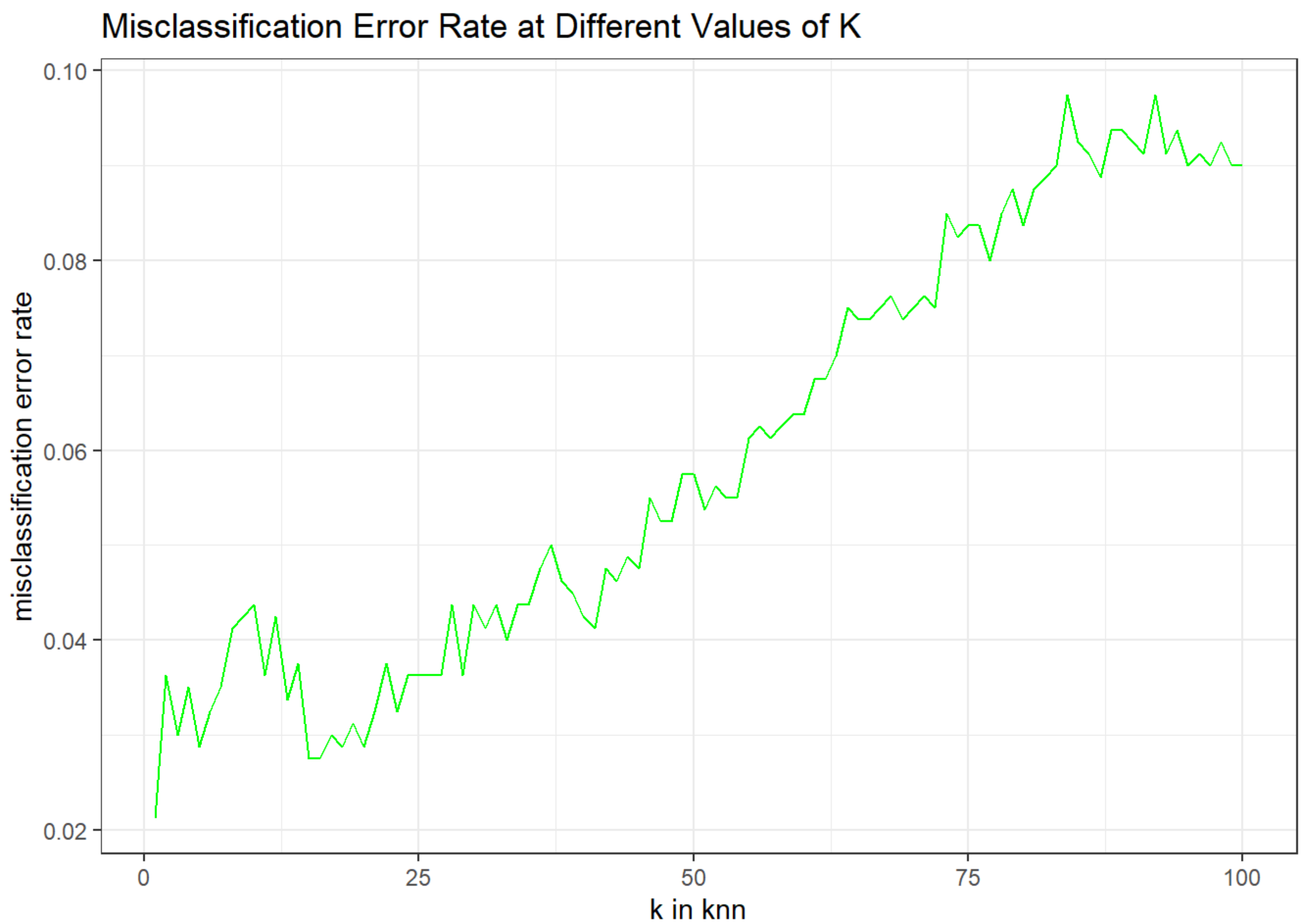
```
summary_table_k_min <- summary_table_k[-1,]
which(summary_table_k_min == min(summary_table_k_min[,4]), arr.ind=TRUE)
```

```
##    row col
## 16  15   4
```

Ignoring k=1, minimum cross-validation error is where k = 15. So our best model is KNN where k = 15.

2b. Plot misclassification error rate at different values of k.

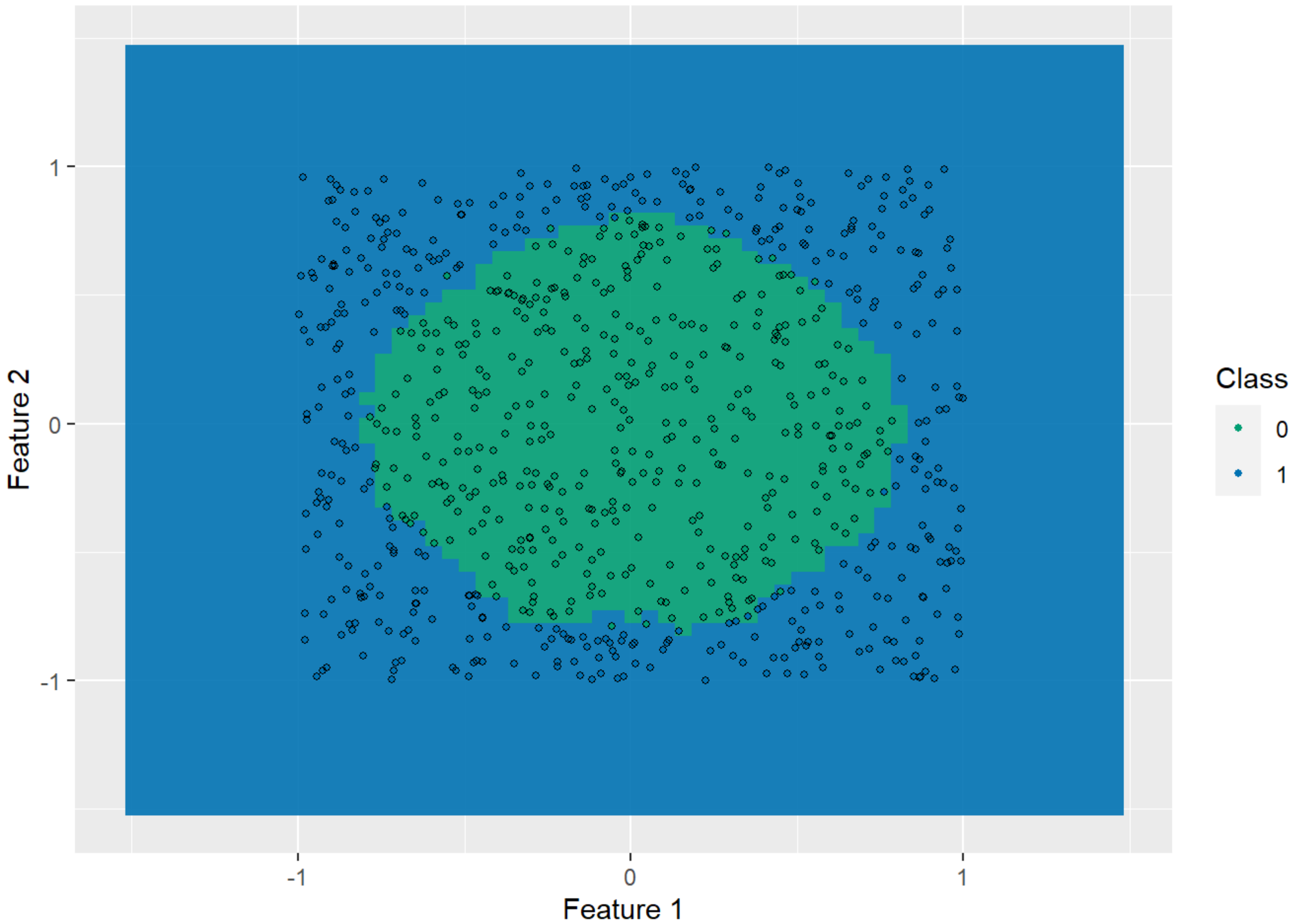
```
library(ggplot2)
ggplot(summary_table_k, aes(x = sum_k)) +
geom_line(aes(y = mean_error), color = "green") + labs(x = "k in knn", y = "misclassification
error rate", title = "Misclassification Error Rate at Different Values of K") + theme_bw()
```



2c. Plot the decision boundary for your classifier using the function at the top code block, `plot_decision_boundary()`. Make sure you load this function into memory before trying to use it.

From 2a, the best model is where k = 15.

```
library(ggplot2)
plot_decision_boundary <- function(tr.x, tr.y, pred_grid, grid)
{
cl <- ifelse(tr.Y == 1, "1", "0")
dataf <- data.frame(grid, prob = as.numeric(pred_grid), class = ifelse(pred_grid==2, "1",
"0"))
col <- c("#009E73", "#0072B2")
plot <- ggplot(dataf) + geom_raster(aes(x=X.1, y=X.2, fill=prob), alpha=.9, data=dataf) +
geom_point(aes(x=X.1, y=X.2, color=class), size=1,
data=data.frame(X.1=tr.X[,1], X.2=tr.X[,2], class=cl)) +
geom_point(aes(x=X.1, y=X.2), size=1, shape=1,
data=data.frame(X.1=tr.X[,1], X.2=tr.X[,2], class=cl)) +
scale_colour_manual(values=col, name="Class") +
scale_fill_gradientn(colors=col[c(1,2)], limits=c(0,1), guide = FALSE) + xlab("Feature 1") +
ylab("Feature 2")
return(plot)
}

tr.X <- train[,1:2]
tr.Y <- train[,3]
te.X <- test[,1:2]
te.Y <- test[,3]
grid <- expand.grid(X.1=seq(min(tr.X[,1]-0.5), max(tr.X[,1]+0.5), by=0.05), X.2=seq(min(tr.X
[,2]-0.5), max(tr.X[,2]+0.5), by=0.05))
y_pred15 <- knn(tr.X, te.X, tr.Y, k =15, prob = TRUE)
pred_grid <- as.numeric(knn(tr.X, grid, tr.Y, k=15, prob=TRUE)) - 1
plot_decision_boundary(tr.X, tr.Y, pred_grid, grid)
```



3. Performance measures for classification

Recall the `Caravan` data from the week 2 lab (part of the `ISLR` package). Train a KNN model with k=2 using all the predictors in the dataset and the outcome `Purchase`. Create a confusion matrix with the test set predictions and the actual values of `Purchase`. Using the values of the confusion matrix, calculate precision, recall, and F1. (Note that `Yes` is the positive class and the confusion matrix may be differently oriented than the one presented in class.)

```
library(ISLR)
names(Caravan)
```

```
##  [1] "MOSTYPE"  "MAANTHUI" "MGEMOMV"  "MGEMLEEF" "MOSHOOFD" "MGODRK"
##  [7] "MGODPR"   "MGODOV"   "MGODGE"   "MRELGE"   "MRELSA"   "MRELOV"
## [13] "MFALLEEN" "MFGEKIND" "MFWEKIND" "MOPLHOOG" "MOPLMIDD" "MOPLLAAG"
## [19] "MBERHOOG" "MBERZELF" "MBERBOER" "MBERMIDD" "MBERARBG" "MBERARBO"
## [25] "MSKA"     "MSKB1"    "MSKB2"    "MSKC"     "MSKD"     "MHHUUR"
## [31] "MHKOOP"   "MAUT1"    "MAUT2"    "MAUT0"    "MZFONDS"  "MZPART"
## [37] "MINKM30"  "MINK3045" "MINK4575" "MINK7512" "MINK123M" "MINKGEM"
## [43] "MKOOPKLA" "PWAPART"  "PWABEDR"  "PWALAND"  "PPERSAUT" "PBESAUT"
## [49] "PMOTSCO"  "PVRAAUT"  "PAANHANG" "PTRACTOR" "PWERKT"   "PBROM"
## [55] "PLEVEN"   "PPERSONG" "PGEZONG"  "PWAOREG"  "PBRAND"   "PZEILPL"
## [61] "PPLEZIER" "PFIETS"   "PINBOED"  "PBYSTAND" "AWAPART"  "AWABEDR"
## [67] "AWALAND"  "APERSAUT" "ABESAUT"  "AMOTSCO"  "AVRAAUT"  "AAANHANG"
## [73] "ATRACTOR" "AWERKT"   "ABROM"    "ALEVEN"   "APERSONG" "AGEZONG"
## [79] "AWAOREG"  "ABRAND"   "AZEILPL"  "APLEZIER" "AFIETS"   "AINBOED"
## [85] "ABYSTAND" "Purchase"
```

```
X <- Caravan[,1:85]
Y <- Caravan[,86]
X <- scale(X)
n_test <- floor(nrow(Caravan) * 0.2)
idx <- sample(1:nrow(Caravan), n_test)
tr.X <- X[-idx,]
te.X <- X[idx,]
tr.Y <- Y[-idx]
te.Y <- Y[idx]
set.seed(1)
pred.Y <- knn(tr.X, te.X, tr.Y, k = 2)
matrix <- table(te.Y, pred.Y)
matrix
```

```
##      pred.Y
## te.Y    No  Yes
##   No  1027   70
##   Yes   59    8
```

```
TP <- matrix['Yes', 'Yes']
FP <- matrix['No', 'Yes']
TN <- matrix['No', 'No']
FN <- matrix['Yes', 'No']
Precision = TP/ (TP + FP)
Precision
```

```
## [1] 0.1025641
```

```
Recall = TP/ (TP + FN)
Recall
```

```
## [1] 0.119403
```

```
F1 = 2*((Precision*Recall)/(Precision+Recall))
F1
```

```
## [1] 0.1103448
```