

Handwriting Digit Recognition

Worked with: Rukku Singla

In order to train the data, Random Forest technique has been used. Random Forest is an ensemble of decision trees. The performance is pretty good in ensemble classifiers as compared to individual classifiers. Here, a sample of the train data has been taken, and the decision trees are built on it.

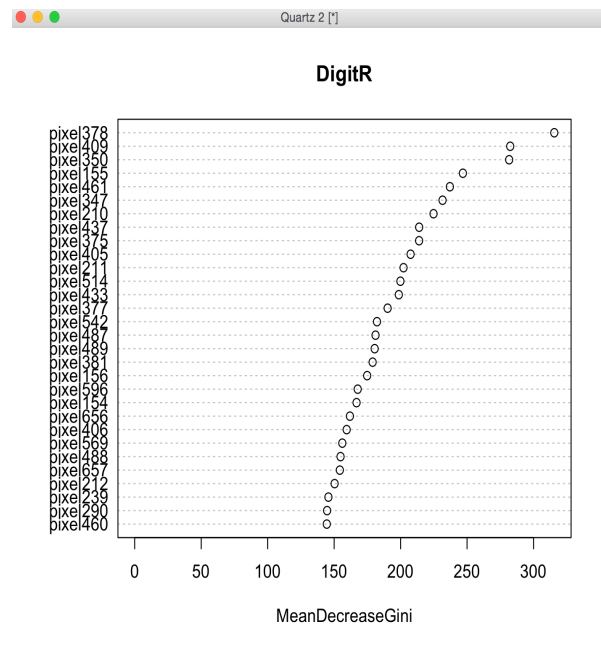
Results:

The Gini Index and the out of bag error have been computed and plotted. For OOB, a dataset different from the sample used for training is used as test data.

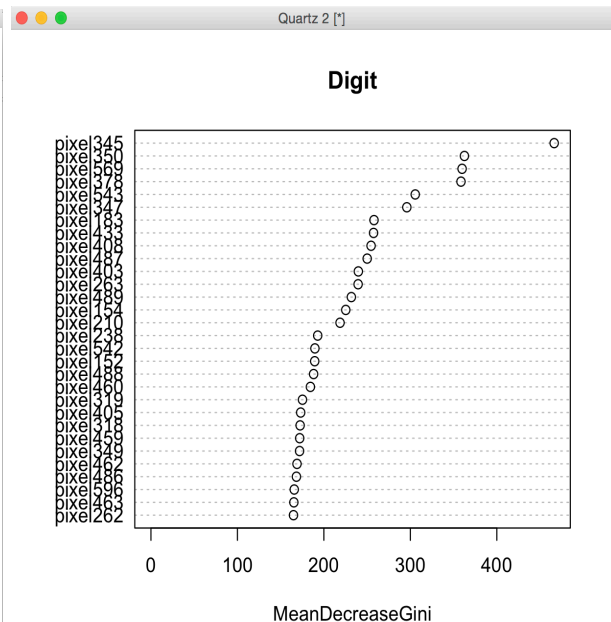
A correlation between the ensemble of classifiers is calculated. The majority vote among the ensemble is used to classify the data.

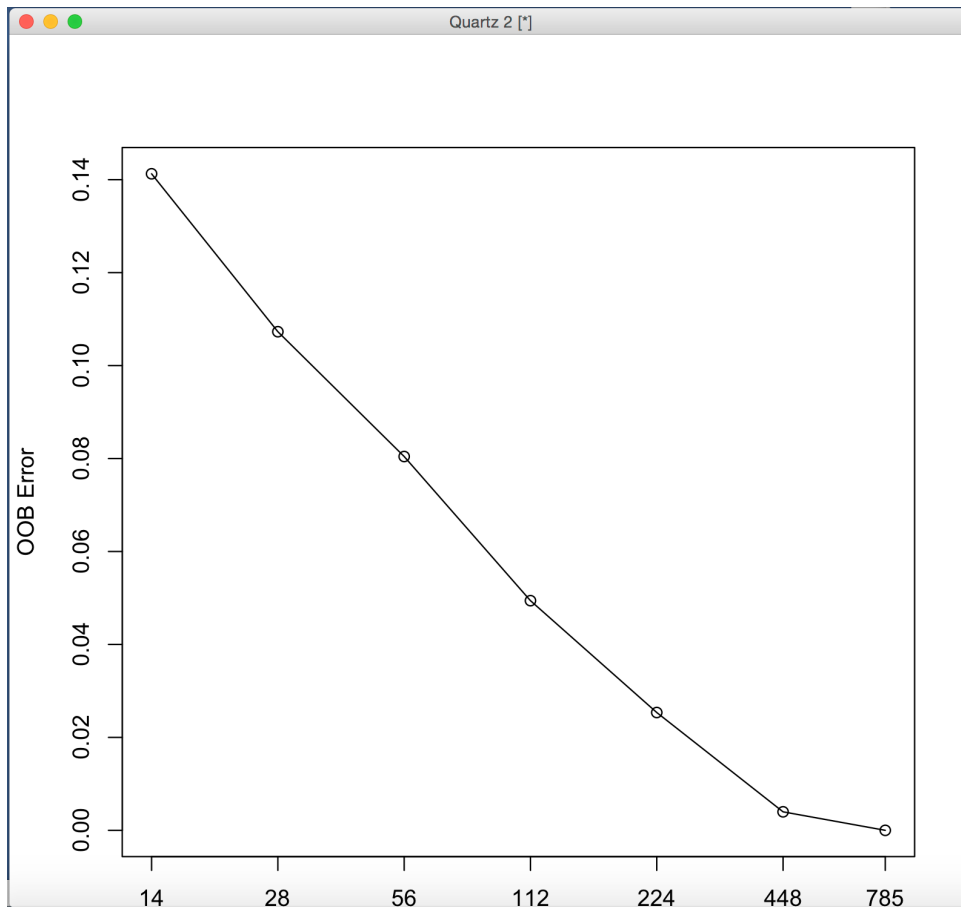
The results of both have been computed and plotted as shown below:

When Ntree=100



When ntree=10





The tune RF is done starting with the default value of mtry, search for the optimal value (with respect to Out-of-Bag error estimate) of mtry for randomForest.
Value of ntree = 100. Accuracy for **ntree=100** is **95.7619%**

```

~/Documents/MNIST_EC 2
type: r11naskj to see new features/changes/vlog r11naskj
> train$label = factor(train$label)
> test$label = factor(test$label)
> DigitR = randomForest(label ~ ., data=train, ntree=100, do.trace=TRUE, mtry=60, nodesize=10)
ntree  OOB  1  2  3  4  5  6  7  8  9  10
1: 20.18% 11.67% 6.94% 22.12% 29.83% 22.10% 30.51% 15.79% 14.79% 26.38% 24.28%
2: 20.22% 9.63% 7.43% 22.88% 28.66% 22.41% 29.20% 16.42% 14.86% 28.53% 24.77%
3: 19.21% 9.85% 7.18% 21.70% 27.00% 21.51% 26.93% 15.84% 12.91% 28.59% 23.17%
4: 18.43% 11.64% 6.60% 20.35% 24.88% 20.17% 25.63% 14.96% 13.39% 26.40% 22.73%
5: 17.65% 11.26% 6.60% 20.55% 22.93% 18.70% 23.70% 13.87% 14.02% 26.07% 20.93%
6: 16.72% 9.72% 6.01% 19.42% 21.88% 17.64% 22.85% 13.43% 13.28% 24.28% 20.75%
7: 15.63% 9.19% 5.11% 17.72% 20.79% 16.40% 22.08% 11.48% 12.33% 23.18% 19.88%
8: 14.87% 8.96% 5.11% 16.87% 20.07% 15.91% 21.03% 11.20% 11.45% 21.85% 18.08%
9: 13.65% 8.23% 4.23% 15.80% 18.61% 14.39% 18.68% 10.14% 10.71% 20.37% 17.04%
10: 12.57% 7.28% 4.00% 13.91% 17.94% 13.24% 17.05% 9.21% 10.38% 19.00% 15.15%
11: 11.94% 6.31% 3.69% 14.38% 16.50% 12.54% 16.35% 8.31% 9.64% 18.16% 14.91%
12: 11.40% 6.18% 3.25% 13.28% 15.92% 11.87% 15.42% 8.60% 9.21% 17.51% 14.12%
13: 10.62% 5.37% 3.18% 12.66% 14.59% 10.73% 15.11% 7.31% 9.14% 16.39% 13.03%
14: 10.04% 5.09% 3.18% 12.28% 14.10% 9.85% 14.74% 6.71% 8.22% 14.85% 12.57%
15: 9.49% 4.67% 2.65% 11.39% 13.27% 9.59% 13.69% 6.57% 7.89% 14.66% 11.75%
16: 9.21% 4.33% 2.65% 11.08% 12.88% 9.03% 13.01% 6.67% 7.37% 14.38% 11.89%
17: 8.70% 4.36% 2.59% 10.53% 12.29% 8.39% 12.51% 5.46% 6.98% 13.71% 11.30%
18: 8.41% 3.98% 2.53% 9.75% 12.05% 7.83% 12.05% 5.63% 7.08% 13.29% 10.96%
19: 8.19% 3.70% 2.56% 9.54% 11.92% 7.82% 11.52% 5.35% 6.91% 12.83% 10.71%
20: 7.97% 3.60% 2.56% 9.23% 11.72% 7.51% 11.63% 4.87% 6.56% 12.66% 10.30%
21: 7.80% 3.49% 2.50% 8.96% 11.79% 7.37% 11.11% 5.08% 6.13% 12.24% 10.20%
22: 7.47% 3.35% 2.44% 8.69% 10.80% 7.12% 10.88% 4.90% 5.97% 11.43% 9.99%
23: 7.44% 3.28% 2.47% 8.58% 11.13% 7.19% 10.43% 4.77% 6.10% 10.94% 10.33%
24: 7.21% 3.01% 2.53% 8.52% 10.67% 6.91% 9.71% 4.80% 5.81% 11.18% 9.69%
25: 7.13% 3.01% 2.41% 8.11% 10.60% 6.70% 9.83% 4.56% 6.04% 10.79% 10.03%
26: 6.91% 2.84% 2.41% 7.76% 10.01% 6.70% 9.41% 4.49% 6.13% 10.58% 9.52%
27: 6.79% 2.63% 2.35% 8.00% 9.78% 6.53% 9.34% 4.28% 6.10% 10.41% 9.21%
28: 6.59% 2.63% 2.23% 7.32% 9.68% 6.49% 9.34% 4.28% 5.78% 9.67% 9.17%
29: 6.57% 2.46% 2.23% 7.56% 9.78% 6.42% 9.00% 4.21% 5.94% 9.63% 9.17%
30: 6.40% 2.39% 2.23% 7.66% 9.49% 6.25% 9.00% 4.07% 5.65% 9.39% 8.59%
31: 6.29% 2.25% 2.23% 7.15% 9.32% 6.11% 8.70% 3.80% 5.81% 9.35% 8.83%
32: 6.24% 2.35% 2.17% 7.25% 9.59% 5.96% 8.43% 3.69% 5.62% 9.18% 8.73%
33: 6.12% 2.21% 2.17% 6.94% 9.32% 5.90% 8.21% 3.73% 6.00% 8.72% 8.49%
34: 5.94% 2.28% 2.13% 6.63% 9.03% 5.44% 8.13% 3.69% 5.52% 8.58% 8.53%
35: 5.99% 2.35% 2.13% 6.70% 8.93% 5.72% 8.40% 3.59% 5.58% 8.61% 8.49%
36: 5.90% 2.35% 2.17% 6.36% 9.09% 5.61% 7.91% 3.63% 5.32% 8.63% 8.49%
37: 5.74% 2.14% 2.17% 6.16% 8.63% 5.61% 7.04% 3.47% 5.37% 8.17% 8.43%

```

21:	7.80%	3.49%	2.50%	8.96%	11.79%	7.37%	11.11%	5.08%	6.13%	12.24%	10.20%
22:	7.47%	3.35%	2.44%	8.69%	10.80%	7.12%	10.88%	4.90%	5.97%	11.43%	9.99%
23:	7.44%	3.28%	2.47%	8.58%	11.13%	7.19%	10.43%	4.77%	6.10%	10.94%	10.33%
24:	7.21%	3.01%	2.53%	8.52%	10.67%	6.91%	9.71%	4.80%	5.81%	11.18%	9.69%
25:	7.13%	3.01%	2.41%	8.11%	10.60%	6.70%	9.83%	4.56%	6.04%	10.79%	10.03%
26:	6.91%	2.84%	2.41%	7.76%	10.01%	6.70%	9.41%	4.49%	6.13%	10.58%	9.52%
27:	6.79%	2.63%	2.35%	8.00%	9.78%	6.53%	9.34%	4.28%	6.10%	10.41%	9.21%
28:	6.59%	2.63%	2.23%	7.32%	9.68%	6.49%	9.34%	4.28%	5.78%	9.67%	9.17%
29:	6.57%	2.46%	2.23%	7.56%	9.78%	6.42%	9.00%	4.21%	5.94%	9.63%	9.17%
30:	6.40%	2.39%	2.23%	7.66%	9.49%	6.25%	9.00%	4.07%	5.65%	9.39%	8.59%
31:	6.29%	2.25%	2.23%	7.15%	9.32%	6.11%	8.70%	3.80%	5.81%	9.35%	8.83%
32:	6.24%	2.35%	2.17%	7.25%	9.59%	5.96%	8.43%	3.69%	5.62%	9.18%	8.73%
33:	6.12%	2.21%	2.17%	6.94%	9.32%	5.96%	8.21%	3.73%	6.00%	8.72%	8.49%
34:	5.94%	2.28%	2.13%	6.63%	9.03%	5.44%	8.13%	3.69%	5.52%	8.58%	8.53%
35:	5.99%	2.35%	2.13%	6.70%	8.93%	5.72%	8.40%	3.59%	5.58%	8.61%	8.49%
36:	5.90%	2.35%	2.17%	6.36%	9.09%	5.61%	7.91%	3.63%	5.32%	8.65%	8.49%
37:	5.74%	2.14%	2.17%	6.16%	8.63%	5.51%	7.94%	3.42%	5.32%	8.12%	8.53%
38:	5.67%	2.21%	2.26%	5.88%	8.47%	5.51%	7.76%	3.38%	5.32%	8.12%	8.25%
39:	5.60%	2.21%	2.13%	5.92%	8.50%	5.23%	7.72%	3.42%	5.32%	7.84%	8.15%
40:	5.48%	2.28%	2.20%	5.68%	7.94%	5.16%	7.64%	3.07%	5.26%	7.67%	8.15%
41:	5.45%	2.17%	2.17%	5.16%	8.47%	5.16%	7.53%	3.21%	5.19%	7.59%	8.05%
42:	5.38%	2.21%	2.10%	5.71%	8.08%	5.05%	7.64%	3.07%	5.23%	7.35%	7.88%
43:	5.35%	2.04%	2.13%	5.44%	8.27%	5.02%	7.83%	2.97%	5.16%	7.21%	7.95%
44:	5.30%	2.01%	2.17%	5.27%	8.60%	4.63%	7.45%	3.00%	5.10%	7.14%	8.02%
45:	5.21%	1.97%	2.17%	5.27%	8.11%	4.53%	7.42%	2.94%	5.03%	7.07%	8.05%
46:	5.18%	2.01%	2.04%	5.51%	8.01%	4.74%	7.04%	2.97%	5.13%	6.75%	8.02%
47:	5.23%	2.18%	2.01%	5.68%	7.94%	4.81%	7.08%	3.07%	5.16%	6.96%	7.81%
48:	5.15%	2.01%	1.98%	5.64%	7.68%	4.81%	7.08%	3.07%	5.13%	6.75%	7.81%
49:	5.18%	1.94%	2.01%	5.85%	7.88%	4.70%	7.23%	2.90%	5.16%	6.86%	7.67%
50:	5.15%	1.90%	1.98%	5.78%	7.68%	4.88%	7.23%	3.11%	4.93%	6.89%	7.61%
51:	5.07%	1.97%	1.98%	5.61%	7.81%	4.74%	7.12%	2.94%	4.84%	6.36%	7.74%
52:	5.01%	1.94%	1.98%	5.47%	7.49%	4.70%	6.97%	3.14%	4.90%	6.29%	7.64%
53:	5.02%	1.87%	1.98%	5.71%	7.52%	4.74%	6.70%	3.11%	4.84%	6.65%	7.47%
54:	5.00%	1.83%	2.07%	5.47%	7.52%	4.53%	7.08%	3.00%	4.97%	6.47%	7.44%
55:	4.99%	1.83%	1.95%	5.44%	7.62%	4.81%	7.04%	2.94%	4.64%	6.43%	7.61%
56:	4.99%	1.90%	1.95%	5.44%	7.58%	4.77%	6.85%	3.00%	4.67%	6.68%	7.47%
57:	4.95%	1.80%	1.95%	5.16%	7.58%	4.70%	6.89%	3.00%	4.71%	6.50%	7.57%
58:	4.91%	1.87%	1.83%	5.37%	7.78%	4.74%	6.66%	2.87%	4.64%	6.15%	7.57%
59:	4.92%	1.97%	1.95%	5.23%	7.88%	4.49%	6.59%	2.90%	4.64%	6.43%	7.50%
60:	4.96%	1.94%	1.86%	5.37%	7.65%	4.53%	6.85%	3.04%	4.74%	6.43%	7.57%
61:	4.86%	1.90%	1.95%	5.23%	7.42%	4.60%	6.63%	2.83%	4.71%	6.47%	7.23%
62:	4.89%	1.90%	1.83%	5.30%	7.42%	4.60%	6.93%	2.83%	4.67%	6.43%	7.40%
63:	4.85%	1.97%	1.83%	5.23%	7.49%	4.63%	6.70%	2.83%	4.71%	6.36%	7.16%
64:	4.79%	1.94%	1.86%	5.06%	7.55%	4.74%	6.40%	2.73%	4.71%	6.19%	7.09%
65:	4.74%	1.94%	1.89%	5.06%	7.45%	4.42%	6.36%	2.69%	4.71%	6.22%	7.03%
66:	4.70%	1.97%	1.80%	4.99%	7.26%	4.35%	6.29%	2.66%	4.67%	6.29%	7.13%
67:	4.72%	1.97%	1.80%	5.13%	7.22%	4.39%	6.36%	2.80%	4.67%	6.40%	6.86%
68:	4.74%	1.97%	1.80%	5.06%	7.29%	4.42%	6.29%	2.83%	4.74%	6.19%	7.20%
69:	4.70%	1.94%	1.74%	4.96%	7.19%	4.42%	6.21%	2.76%	4.74%	6.26%	7.13%
70:	4.68%	1.97%	1.77%	5.16%	7.16%	4.32%	6.06%	2.59%	4.74%	6.40%	6.96%
71:	4.65%	2.01%	1.74%	4.99%	7.09%	4.46%	6.14%	2.49%	4.87%	6.26%	6.82%
72:	4.64%	2.11%	1.77%	5.10%	7.09%	4.32%	5.95%	2.49%	4.67%	6.29%	6.92%
73:	4.66%	1.94%	1.74%	4.96%	7.09%	4.39%	6.33%	2.38%	4.84%	6.36%	6.96%
74:	4.64%	2.01%	1.74%	4.99%	7.12%	4.42%	6.17%	2.38%	4.84%	6.05%	7.03%
75:	4.59%	1.94%	1.71%	4.92%	7.26%	4.39%	5.99%	2.35%	4.87%	5.91%	6.89%
76:	4.57%	1.97%	1.77%	5.03%	7.09%	4.46%	5.95%	2.42%	4.64%	5.84%	6.86%
77:	4.58%	1.94%	1.71%	4.89%	7.09%	4.39%	6.02%	2.38%	4.90%	5.87%	6.92%
78:	4.54%	2.04%	1.77%	4.89%	7.03%	4.39%	5.95%	2.49%	4.77%	5.63%	6.79%
79:	4.54%	1.97%	1.65%	4.82%	7.16%	4.35%	5.87%	2.42%	4.80%	5.56%	7.09%
80:	4.49%	2.04%	1.77%	4.86%	6.89%	4.32%	5.87%	2.28%	4.77%	5.41%	6.96%
81:	4.53%	2.01%	1.74%	4.72%	7.12%	4.28%	5.80%	2.49%	4.84%	5.59%	6.99%
82:	4.48%	2.07%	1.62%	4.72%	6.96%	4.18%	5.65%	2.52%	4.87%	5.63%	6.92%
83:	4.49%	2.01%	1.83%	4.82%	7.03%	4.04%	5.87%	2.42%	4.67%	5.66%	6.82%
84:	4.48%	2.07%	1.74%	4.72%	7.06%	4.07%	5.84%	2.49%	4.84%	5.52%	6.72%
85:	4.48%	1.94%	1.68%	4.62%	6.89%	4.11%	5.91%	2.49%	4.97%	5.63%	6.86%
86:	4.41%	1.94%	1.65%	4.55%	6.86%	4.00%	5.76%	2.31%	4.93%	5.52%	6.86%
87:	4.44%	2.01%	1.65%	4.55%	6.96%	4.00%	5.80%	2.45%	4.84%	5.77%	6.72%
88:	4.50%	1.90%	1.83%	4.65%	6.86%	4.14%	5.91%	2.52%	5.00%	5.73%	6.75%
89:	4.45%	1.90%	1.74%	4.72%	6.93%	4.00%	5.84%	2.42%	4.87%	5.80%	6.55%
90:	4.46%	1.90%	1.77%	4.65%	7.03%	4.18%	5.53%	2.35%	5.03%	5.77%	6.65%
91:	4.44%	1.87%	1.68%	4.62%	6.93%	4.07%	5.76%	2.45%	4.90%	5.77%	6.68%
92:	4.44%	1.94%	1.74%	4.62%	6.76%	4.07%	5.69%	2.38%	4.90%	5.80%	6.79%
93:	4.42%	1.90%	1.68%	4.62%	6.80%	4.11%	5.80%	2.31%	4.80%	5.73%	6.79%
94:	4.41%	1.90%	1.71%	4.51%	6.76%	4.14%	5.76%	2.56%	4.67%	5.63%	6.75%
95:	4.40%	1.90%	1.68%	4.55%	6.57%	4.07%	5.46%	2.49%	5.03%	5.73%	6.86%
96:	4.36%	1.90%	1.71%	4.41%	6.70%	4.07%	5.57%	2.45%	4.80%	5.66%	6.62%
97:	4.41%	1.97%	1.65%	4.55%	6.57%	4.11%	5.76%	2.52%	5.10%	5.49%	6.72%
98:	4.33%	1.97%	1.62%	4.51%	6.53%	4.07%	5.50%	2.38%	4.77%	5.49%	6.79%
99:	4.36%	1.94%	1.71%	4.69%	6.70%	4.14%	5.35%	2.45%	4.71%	5.63%	6.55%
100:	4.32%	2.01%	1.65%	4.38%	6.57%	4.07%	5.31%	2.52%	4.67%	5.66%	6.62%

Value of ntree= 10

> Digit = randomForest(label ~ ., data=train, ntree=10, do.trace=TRUE,mtry=60,nodesize=10)

ntree	00B	1	2	3	4	5	6	7	8	9	10
1:	20.40%	12.56%	5.40%	26.62%	27.91%	21.18%	27.20%	20.70%	13.66%	26.89%	24.17%
2:	20.28%	12.95%	5.43%	24.61%	25.38%	21.84%	28.28%	21.09%	13.73%	27.45%	24.84%
3:	19.64%	11.87%	5.80%	22.54%	24.60%	20.59%	27.91%	19.27%	13.26%	28.71%	24.40%
4:	18.92%	11.11%	6.23%	22.24%	24.09%	19.54%	26.73%	17.05%	13.86%	27.97%	22.61%
5:	18.09%	10.69%	5.73%	21.13%	23.35%	17.99%	25.09%	16.14%	13.66%	27.35%	21.90%
6:	16.85%	9.95%	5.39%	18.72%	22.34%	16.83%	24.57%	14.69%	12.41%	25.09%	20.63%
7:	15.84%	9.09%	5.04%	17.57%	20.80%	15.31%	23.00%	14.11%	11.83%	23.60%	19.98%
8:	14.68%	8.07%	4.78%	15.59%	19.21%	14.51%	21.05%	12.88%	10.93%	22.97%	18.62%
9:	13.99%	7.66%	4.44%	15.81%	18.34%	13.31%	20.54%	11.98%	10.91%	20.85%	17.80%
10:	12.97%	6.65%	4.47%	14.37%	17.81%	11.94%	19.13%	10.86%	9.91%	19.84%	16.31%

Accuracy achieved for ntree=10 is 93.49206%

Below is a sample of the prediction:

4 7 10 13 14 15
 4 7 3 1 3 3
 Levels: 0 1 2 3 4 5 6 7 8 9

	0	1	2	3	4	5	6	7	8	9
0	1221	0	1	1	0	1	5	1	8	2
1	0	1380	6	5	2	3	1	3	4	1
2	7	4	1196	3	12	2	3	13	10	3
3	2	0	28	1230	0	14	3	6	18	4
4	3	3	2	0	1175	2	7	1	5	24
5	7	3	1	18	0	1083	10	2	11	4
6	10	4	1	0	3	9	1202	0	12	0
7	0	3	19	0	14	1	0	1258	5	20
8	1	9	3	15	7	8	5	1	1152	18
9	10	3	6	22	16	4	1	15	10	1169

```
> mtry <- tuneRF(train, train$label, ntreeTry = 10, stepFactor = 2, improve = 0.05, trace = TRUE, plot = TRUE, doBest = FALSE)
mtry = 28 OOB error = 10.73%
Searching left ...
mtry = 14 OOB error = 14.13%
-0.3167135 0.05
Searching right ...
mtry = 56 OOB error = 8.04%
0.2502963 0.05
mtry = 112 OOB error = 4.94%
0.3856051 0.05
mtry = 224 OOB error = 2.54%
0.4869796 0.05
mtry = 448 OOB error = 0.4%
0.8428376 0.05
mtry = 785 OOB error = 0%
1 0.05
> mtry <- tuneRF(train, train$label, ntreeTry = 10, stepFactor = 2, improve = 0.05, trace = TRUE, plot = TRUE, doBest = FALSE)
mtry = 28 OOB error = 10.73%
Searching left ...
mtry = 14 OOB error = 14.13%
-0.3167135 0.05
Searching right ...
mtry = 56 OOB error = 8.04%
0.2502963 0.05
mtry = 112 OOB error = 4.94%
0.3856051 0.05
mtry = 224 OOB error = 2.54%
0.4869796 0.05
mtry = 448 OOB error = 0.4%
0.8428376 0.05
mtry = 785 OOB error = 0%
1 0.05
```

References:

<http://www.inside-r.org/packages/cran/randomforest/docs/tuneRF>
<http://deblivingdata.net/>