

HW12 - MYSTERIOUS BOX FOUND ON THE BEACH

Nicole Coria

TRIAL AND ERROR

```
tree1 <- rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, control=rpart.control(minsplit = 200))
rpart.plot::rpart.plot(tree1)
CrossValidation::cross_validate(train, tree1, 3, 0.8)
```

```
[[[[
  accuracy_subset accuracy_all
1      0.6656510      0.6656510
2      0.6614958      0.6614958
3      0.6728532      0.6728532
```

```
tree2 <- rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, control=rpart.control(minsplit = 50))
rpart.plot::rpart.plot(tree2)
CrossValidation::cross_validate(train, tree2, 3, 0.8)
```

```
[[[[
  accuracy_subset accuracy_all
1      0.6664820      0.6664820
2      0.6562327      0.6562327
3      0.6673130      0.6673130
```

- At first I tried a couple of trees that gave me the same accuracy subset and and accuracy all

TRIAL AND ERROR CONTINUED

```
tree3 <- rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, control=rpart.control(minbucket = 100))
rpart.plot::rpart.plot(tree3)
CrossValidation::cross_validate(train, tree3, 3, 0.8)
```

	accuracy_subset	accuracy_all
1	0.6534626	0.6534626
2	0.6567867	0.6567867
3	0.6495845	0.6495845

```
tree4 <- rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, control=rpart.control(minbucket = 100))
rpart.plot::rpart.plot(tree4)
CrossValidation::cross_validate(train, tree4, 3, 0.8)
```

	accuracy_subset	accuracy_all
1	0.6573407	0.6573407
2	0.6470914	0.6470914
3	0.6509695	0.6509695

- At first I tried a couple of trees that gave me the same accuracy subset and and accuracy all

TRIAL AND ERROR CONTINUED

```
tree5 <- rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, cp = 0.0002, minbucket = 50)
rpart.plot::rpart.plot(tree5)
CrossValidation::cross_validate(train, tree5, 3, 0.8)
```

	accuracy_subset	accuracy_all
1	0.6745152	0.6556787
2	0.6789474	0.6603878
3	0.6706371	0.6659280

```
tree <- rpart::rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, cp = 0.0002, minbucket = 50,
maxdepth=30)
rpart.plot::rpart.plot(tree)
CrossValidation::cross_validate(train, tree, 3, 0.8)
```

	accuracy_subset	accuracy_all
1	0.6642659	0.6443213
2	0.6797784	0.6603878
3	0.6700831	0.6537396

- Here I tried using cp which did help my accuracy subset be greater than accuracy all.
- I did try different values for minbucket, but keeping it around 50 seemed best

BEST DECISION TREE

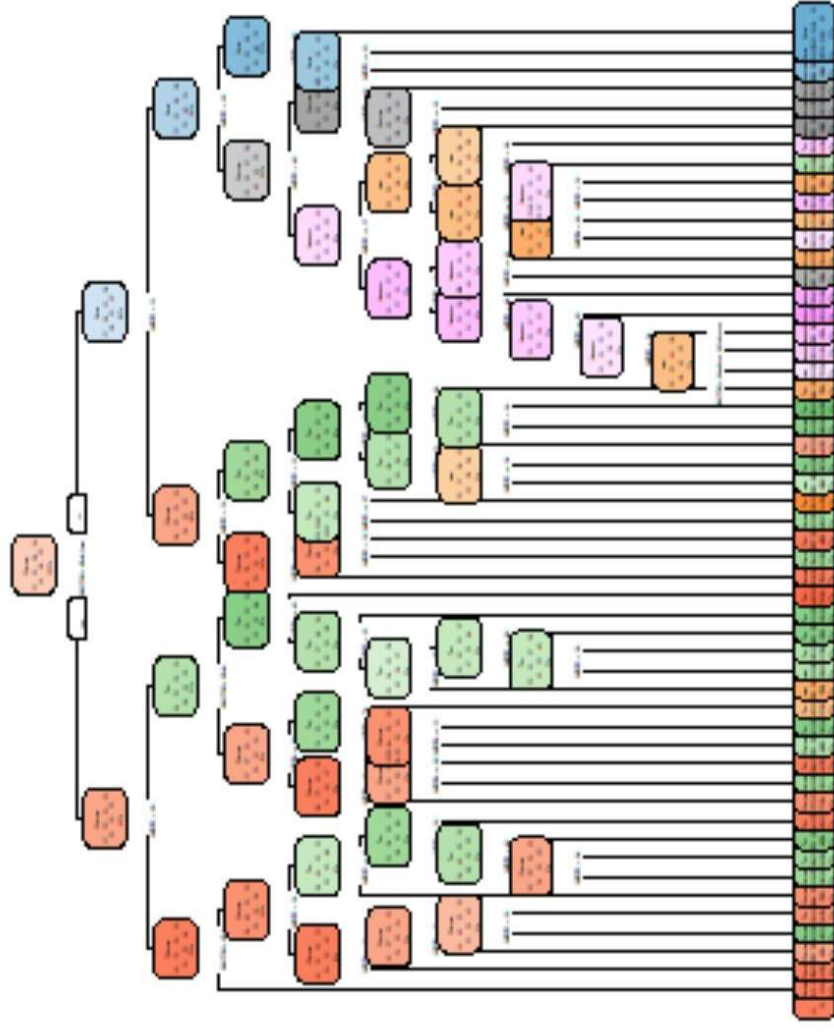
```
tree6 <- rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, cp = 0.0002, minsplit = 50)
```

```
rpart.plot::rpart.plot(tree6)
```

CrossValidation::cross_validate(train, tree6, 3, 0.8)

- After messing around with different controls cp, minsplit and minbucket, this tree was able to give me a higher accuracy subset over the accuracy all

	accuracy_subset	accuracy_all
1	0.6745152	0.6612188
2	0.6950139	0.6703601
3	0.6825485	0.6590028



```
tree6 <- rpart(as.factor(SOUND)~INPUT1+INPUT2+INPUT3+INPUT4+SWITCH, data=train, cp  
= 0.0002, minsplit = 50)
```

```
rpart.plot::rpart.plot(tree6)
```

```
CrossValidation::cross_validate(train, tree6, 3, 0.8)
```