

Tree Model Examples

Some remarks on `rpart()`

$$cp = \alpha / SS_0.$$

xerror is (relative) sum-of-squared errors in tenfold cross-validation.

xstd represents variation in prediction across the ten validation samples.

Simulate some data for tree-fitting examples

```
> set.seed(150)
> x1 <- round(rnorm(100),2); x2 <- round(rnorm(100),2)
> noise <- round(rnorm(100),2)

> y <- x1+noise; dt <- data.frame(cbind(y,x1,x2))

> set.seed(794)
> fit1.a <- rpart(y~x1+x2,data=dt, cp=.001)
> fit1.b <- rpart(y~x1+x2,data=dt, cp=.001)
```

```
> printcp(fit1a)
```

```
Regression tree:
```

```
rpart(formula = y ~ x1 + x2, data = dt, cp = 0.001)
```

```
Root node error: 233.31/100 = 2.3331
```

	CP	nsplit	rel error	xerror	xstd
1	0.3982870	0	1.00000	1.01170	0.136653
2	0.1541180	1	0.60171	0.70437	0.103025
3	0.0384161	2	0.44760	0.49255	0.076145
4	0.0226927	3	0.40918	0.46238	0.069819
5	0.0090841	4	0.38649	0.45191	0.068230
6	0.0060035	6	0.36832	0.49262	0.074234
7	0.0010000	7	0.36231	0.48833	0.072860

```
> fit1.a.pruned <- prune(fit1.a, cp=.02)
```

```
> printcp(fit1.b)
```

```
Regression tree:
```

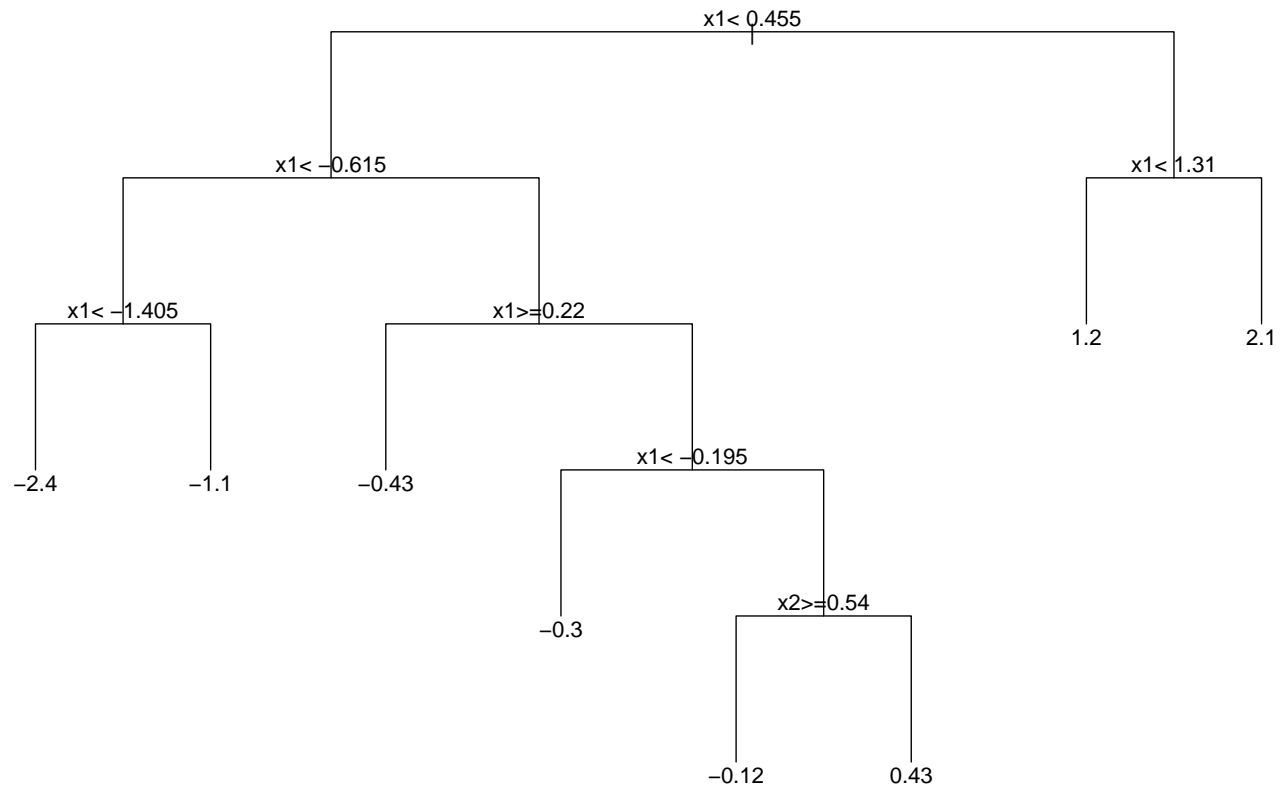
```
rpart(formula = y ~ x1 + x2, data = dt, cp = 0.001)
```

```
Root node error: 233.31/100 = 2.3331
```

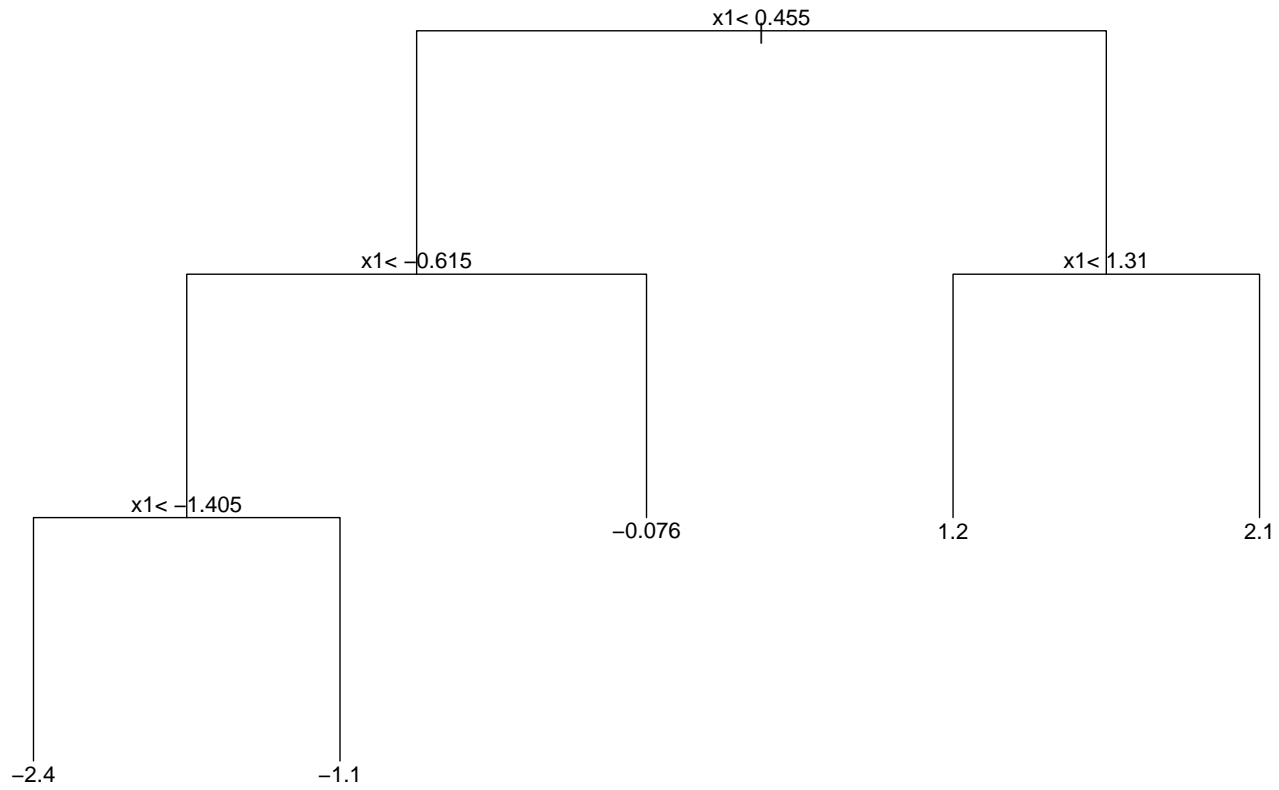
	CP	nsplit	rel error	xerror	xstd
1	0.3982870	0	1.00000	1.01227	0.137338
2	0.1541180	1	0.60171	0.76735	0.112788
3	0.0384161	2	0.44760	0.49054	0.075356
4	0.0226927	3	0.40918	0.49428	0.076309
5	0.0090841	4	0.38649	0.50970	0.077870
6	0.0060035	6	0.36832	0.53877	0.079370
7	0.0010000	7	0.36231	0.53700	0.079119

```
> fit1.b.pruned <- prune(fit1.b, cp=.04)
```

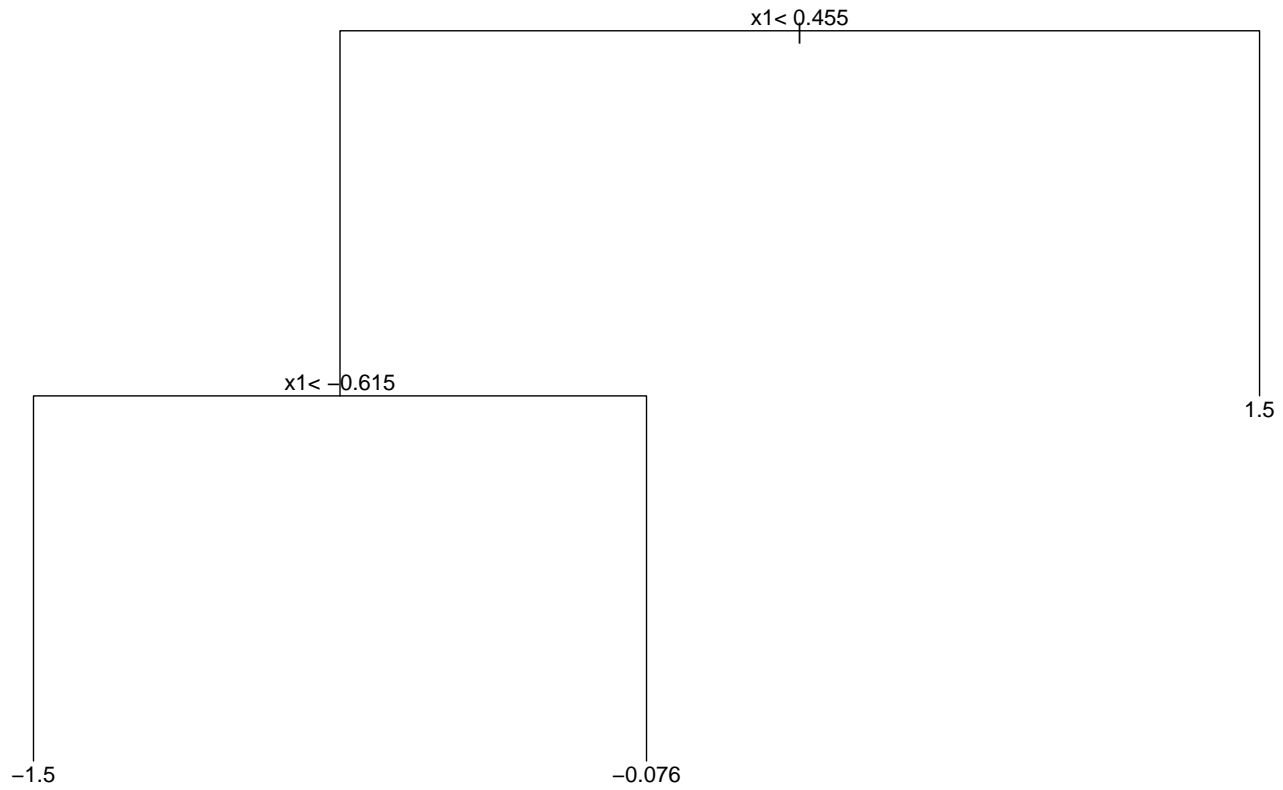
Dataset 1, Unpruned Fit



Dataset 1, First Pruned Fit



Dataset 2, Second Pruned Fit



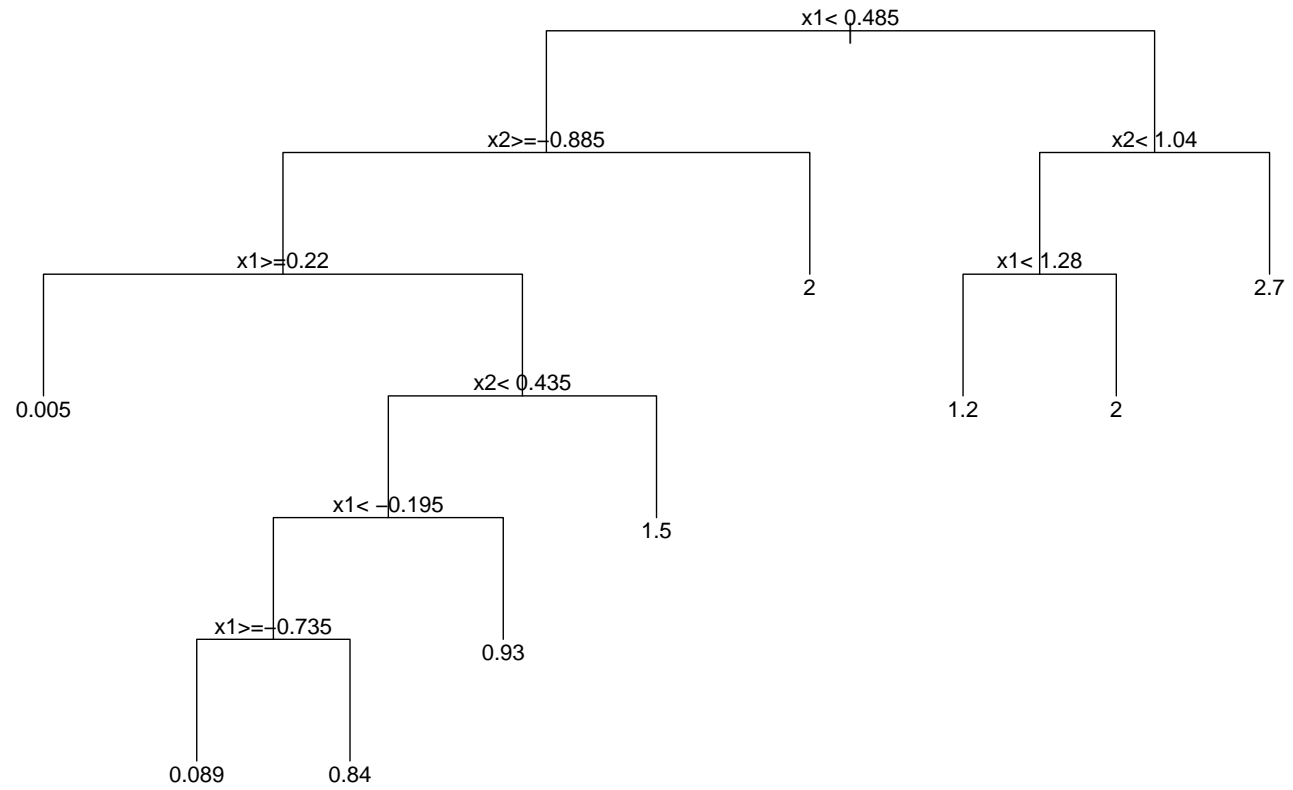
Second example

```
> y <- round(sqrt(x1^2+x2^2)+noise,digits=2)
```

```
> fit2 <- rpart(y~x1+x2,data=dt, cp=.001)
```

	CP	nsplit	rel error	xerror	xstd
1	0.112013	0	1.00000	1.0394	0.15894
2	0.105763	1	0.88799	1.1493	0.16595
3	0.054138	2	0.78222	1.0258	0.14791
4	0.044665	4	0.67395	1.0714	0.14507
5	0.021122	5	0.62928	1.0340	0.15335
6	0.015787	6	0.60816	1.0113	0.14765
7	0.000001	8	0.57659	1.0155	0.15094

Dataset 2, Unpruned Fit



Third Example

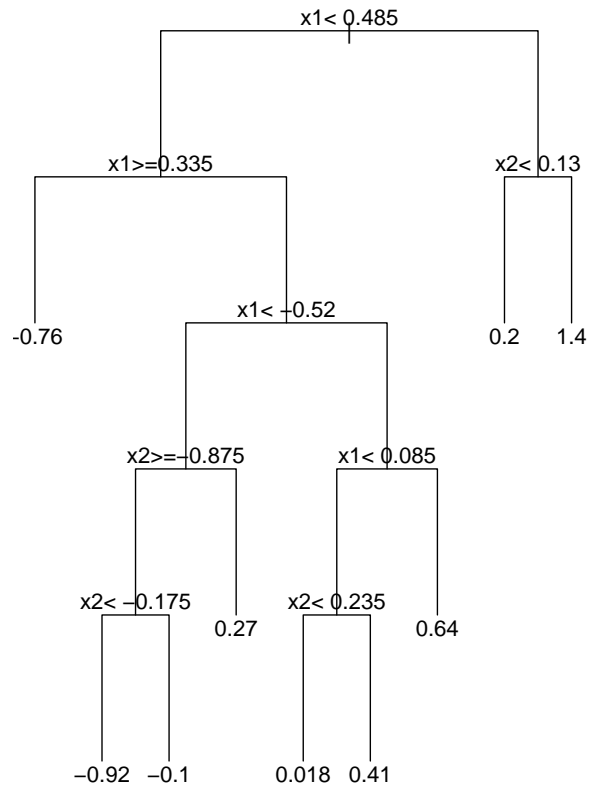
```
> y <- round( as.numeric((x1>0)&(x2>0)) + noise,digits=2)
```

```
> fit3 <- rpart(y~x1+x2,data=dt, cp=.001)
```

	CP	nsplit	rel error	xerror	xstd
1	0.1042658	0	1.00000	1.04412	0.14304
2	0.0799562	1	0.89573	1.08600	0.14176
3	0.0407288	2	0.81578	0.99553	0.14747
4	0.0266698	4	0.73432	1.11006	0.15831
5	0.0148459	6	0.68098	1.13835	0.16006
6	0.0067386	7	0.66614	1.11390	0.15893
7	0.0010000	8	0.65940	1.11464	0.15804

```
> fit3.pruned <- prune(fit3, cp=.05)
```

Dataset 3, Unpruned Fit



Dataset 3, Pruned Fit

