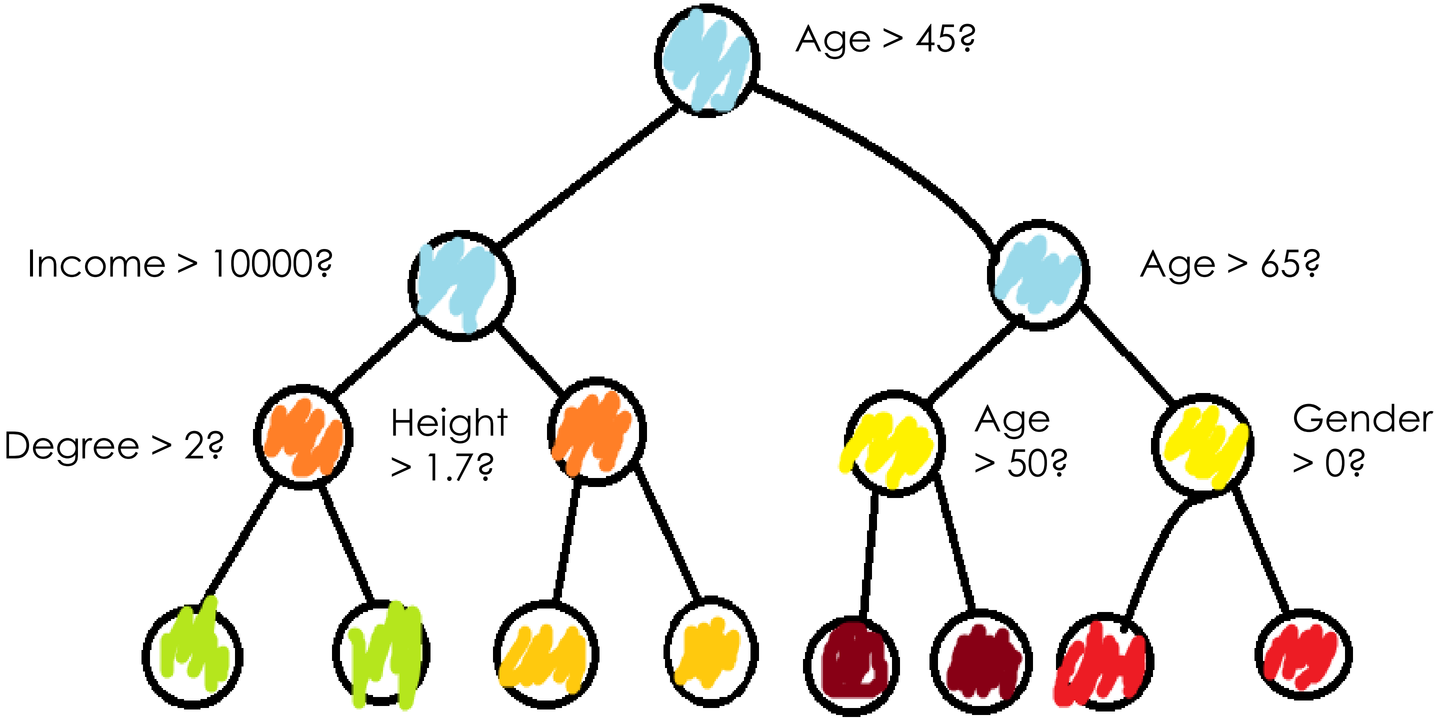
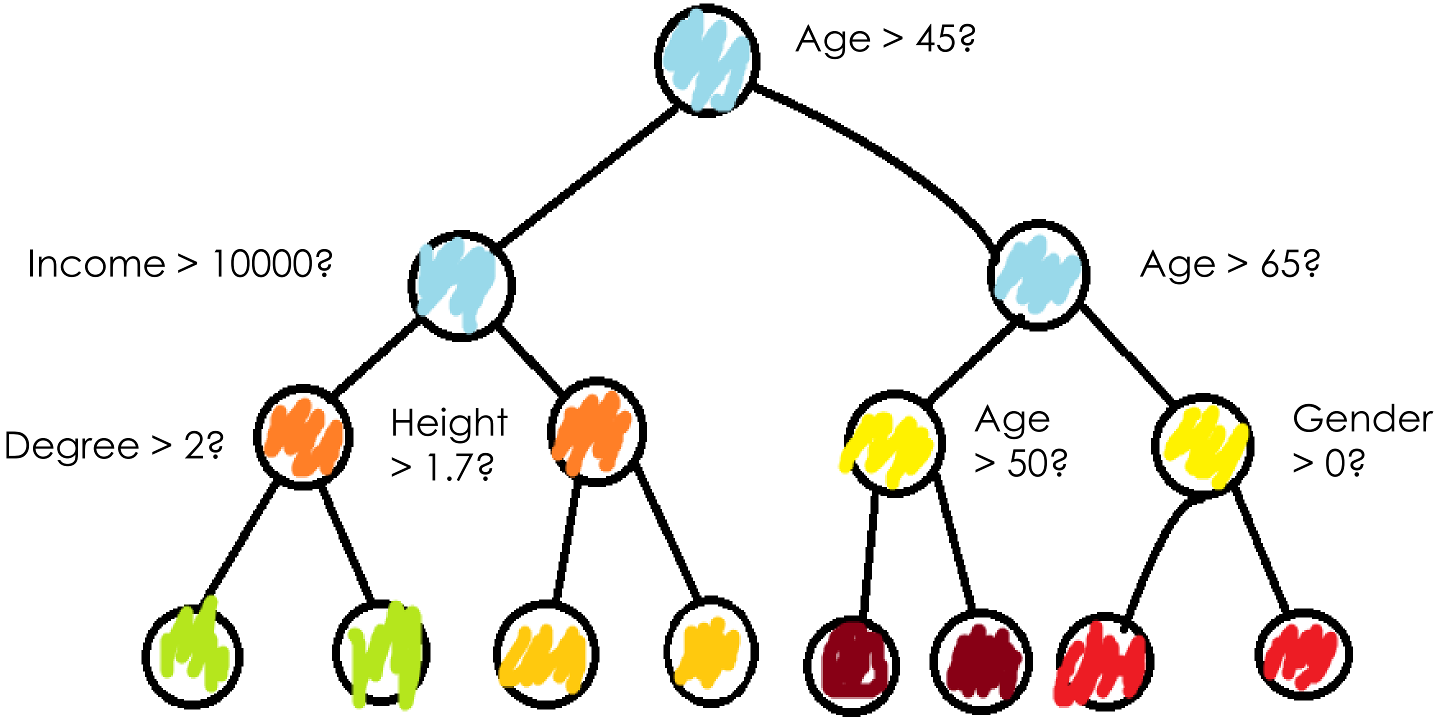
Feature importance

Decision tree: induction



At each note there is an increase in purity.

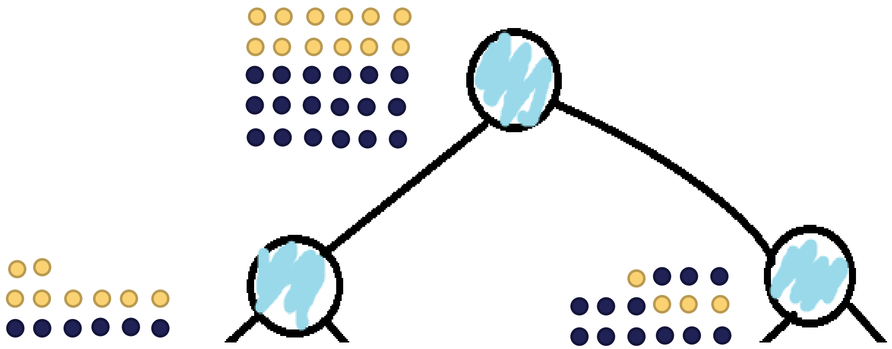
Feature importance



- How often a feature is chosen.

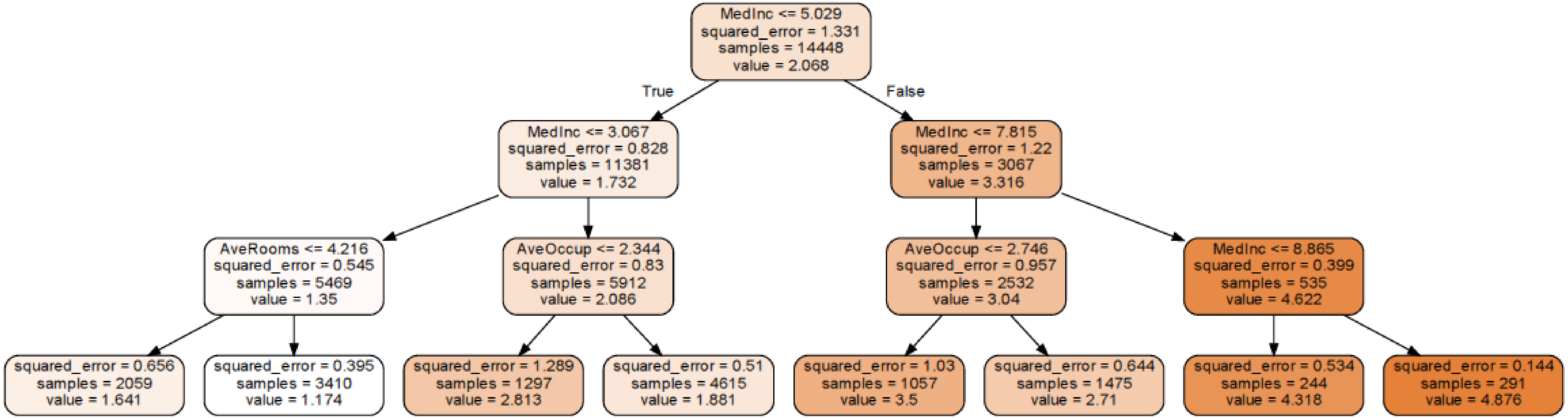
- How big the increase in purity is.

Feature importance

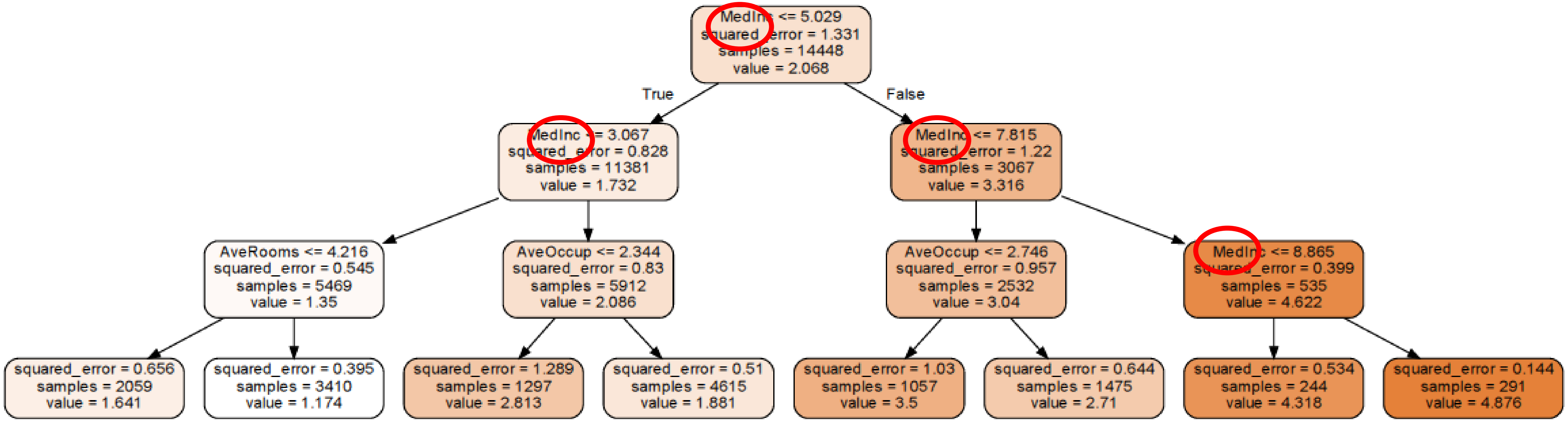


```
importances[node.feature] +=
weighted_n_node_samples * node.impurity –
(left.weighted_n_node_samples * left.impurity +
right.weighted_n_node_samples * right.impurity)
```

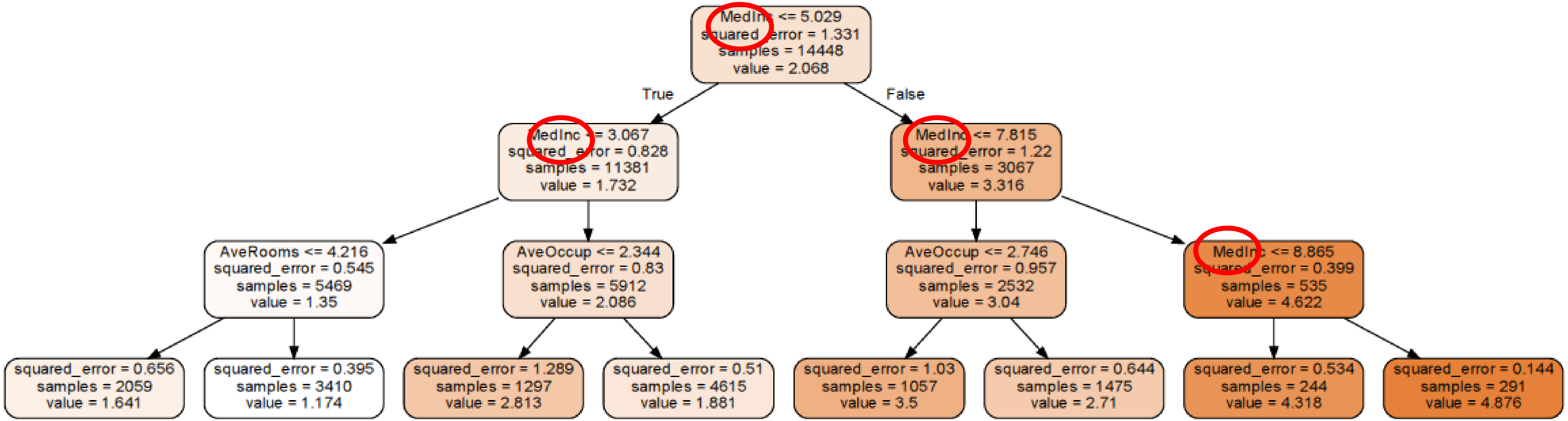
Feature importance



Feature importance



Feature importance



MedInc <= 5.029
squared_error = 1.331
samples = 14448
value = 2.068

True

False

MedInc <= 3.067
squared_error = 0.828
samples = 11381
value = 1.732

MedInc <= 7.815
squared_error = 1.22
samples = 3067
value = 3.316

AveRooms <= 4.216
squared_error = 0.545
samples = 5469
value = 1.35

AveOccup <= 2.344
squared_error = 0.83
samples = 5912
value = 2.086

AveOccup <= 2.746
squared_error = 0.957
samples = 2532
value = 3.04

MedInc <= 8.865
squared_error = 0.399
samples = 535
value = 4.622

squared_error = 0.656
samples = 2059
value = 1.641

squared_error = 0.395
samples = 3410
value = 1.174

squared_error = 1.289
samples = 1297
value = 2.813

squared_error = 0.51
samples = 4615
value = 1.881

squared_error = 1.03
samples = 1057
value = 3.5

squared_error = 0.644
samples = 1475
value = 2.71

squared_error = 0.534
samples = 244
value = 4.318

squared_error = 0.144
samples = 291
value = 4.876

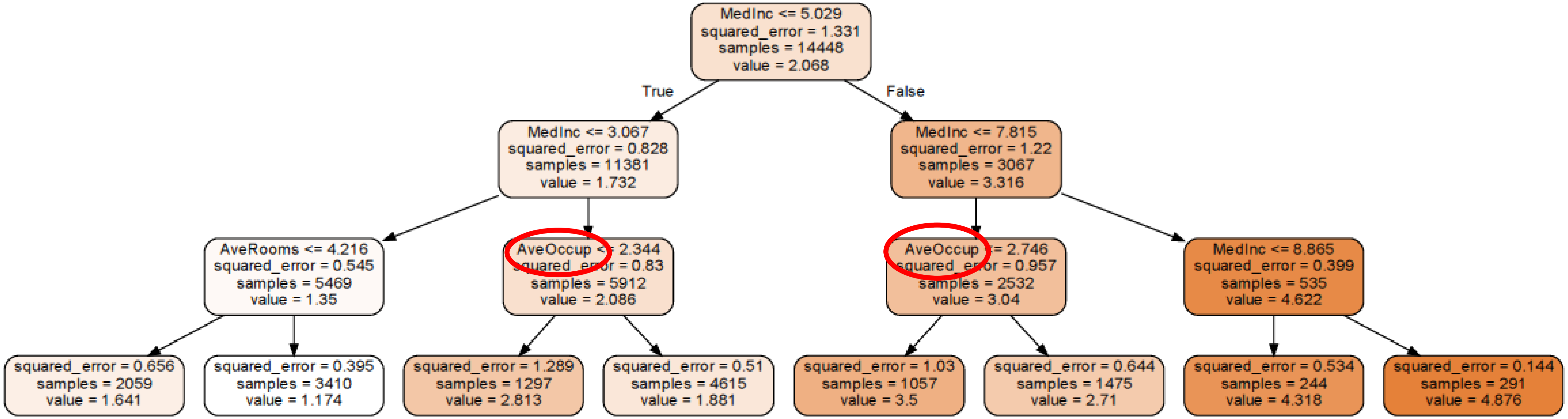
MedInc[node-1] = 1 * 1.331 – (11381/14448 * 0.828 + 3067/14448 * 1.22)

MedInc[node-2] = 1 * 0.828 – (5469/11381 * 0.545 + 5912/11382 * 0.83)

…

MedInc = MedInc[node-1] + MedInc[node-2] … + MedInc[node-n]

Feature importance



MedInc <= 5.029
squared_error = 1.331
samples = 14448
value = 2.068

True

False

MedInc <= 3.067
squared_error = 0.828
samples = 11381
value = 1.732

MedInc <= 7.815
squared_error = 1.22
samples = 3067
value = 3.316

AveRooms <= 4.216
squared_error = 0.545
samples = 5469
value = 1.35

AveOccup <= 2.344
squared_error = 0.83
samples = 5912
value = 2.086

AveOccup <= 2.746
squared_error = 0.957
samples = 2532
value = 3.04

MedInc <= 8.865
squared_error = 0.399
samples = 535
value = 4.622

squared_error = 0.656
samples = 2059
value = 1.641

squared_error = 0.395
samples = 3410
value = 1.174

squared_error = 1.289
samples = 1297
value = 2.813

squared_error = 0.51
samples = 4615
value = 1.881

squared_error = 1.03
samples = 1057
value = 3.5

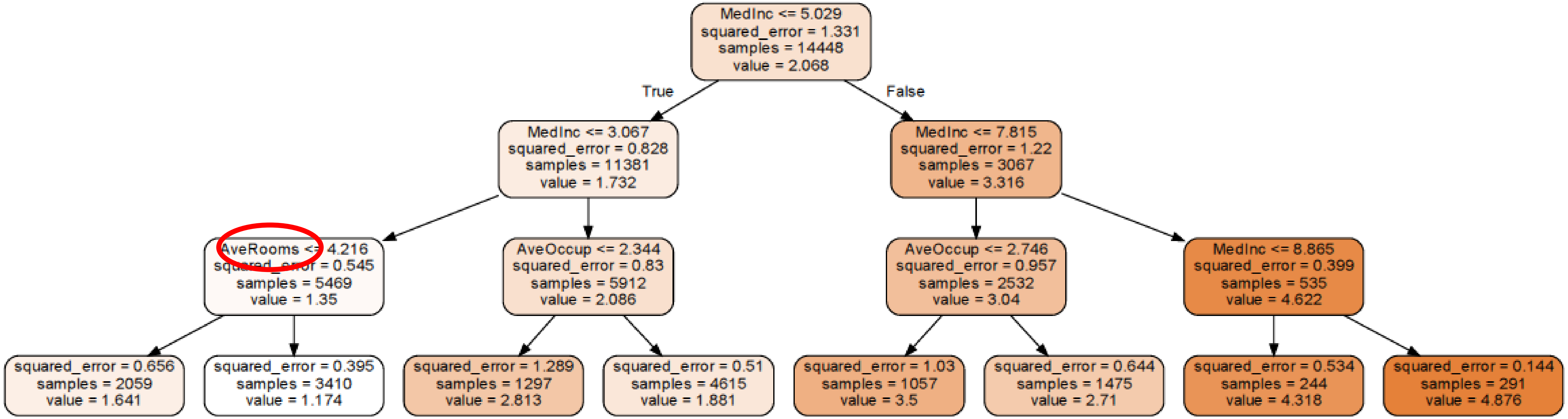
squared_error = 0.644
samples = 1475
value = 2.71

squared_error = 0.534
samples = 244
value = 4.318

squared_error = 0.144
samples = 291
value = 4.876

```
AveOccup = AveOccup[node-1] + AveOccup[node-2]
```

Feature importance



```
AveRooms = AveRooms[node-1]
```

Feature importance

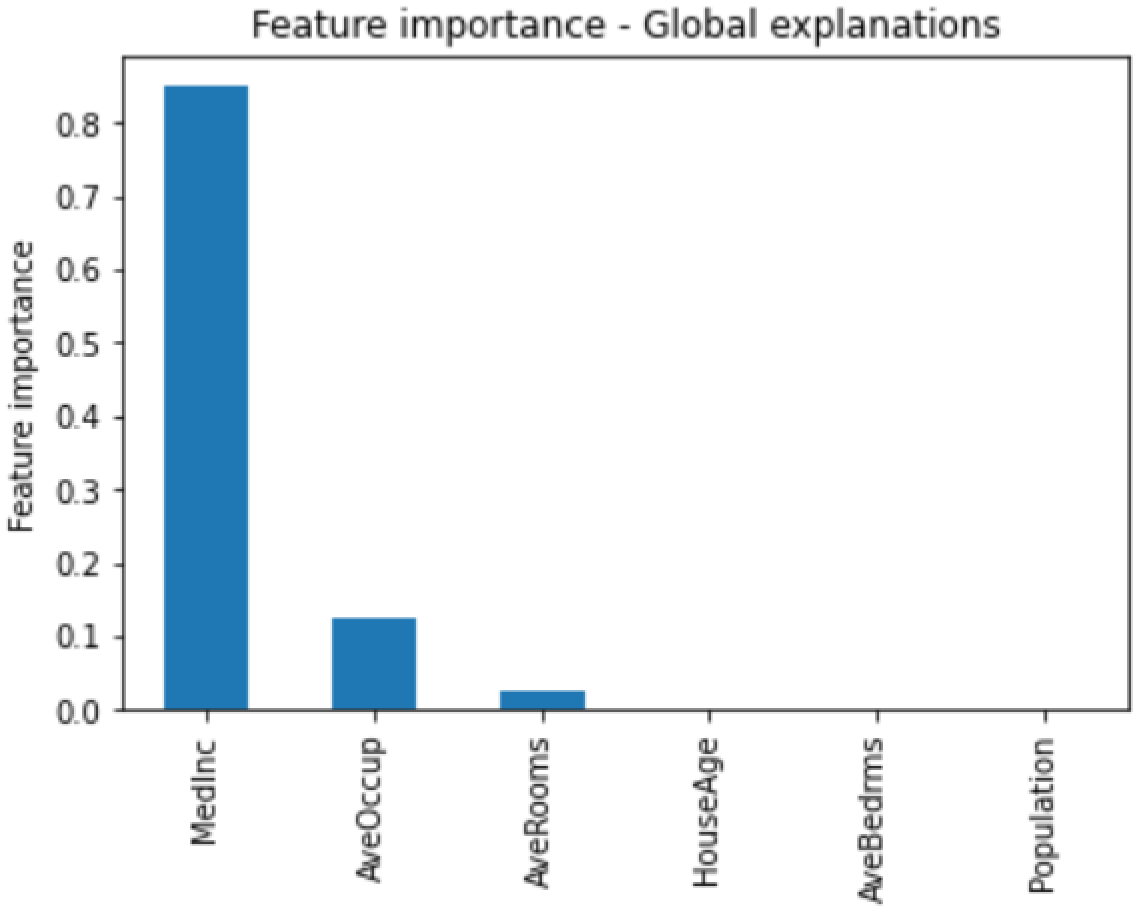
```
Sum(importance) = Importance[MedInc] + Importance[AveOccup] + Importance[AveRooms]
```

```
Importance[MedInc] = Importance[MedInc] / Sum(importance)
```

```
Importance[AveOccup] = Importance[AveOccup] / Sum(importance)
```

```
Importance[AveRooms] = Importance[AveRooms] / Sum(importance)
```

Feature importance



- Features at top nodes have generally greater importance.
  - higher decrease in impurity.

- Features used in multiple nodes are more important.
  - Higher cumulative impurity decrease.

Train In Data

THANK YOU

www.trainindata.com