LEC005 Demand Forecasting

VG441 SS2021

Cong Shi
Industrial & Operations Engineering
University of Michigan

XGBoost

F0 = Initial Model = Taking the mean

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

Height (m)	Favorite Color	Gender	Weight (kg)	F0	PR0
1.6	Blue	Male	88	71.2	16.8
1.6	Green	Female	76	71.2	4.8
1.5	Blue	Female	56	71.2	-15.2
1.8	Red	Male	73	71.2	1.8
1.5	Green	Male	77	71.2	5.8
1.4	Blue	Female	57	71.2	-14.2

Pseudo Residual (PR) = True Value – Predicted Value

XGBoost Tree

Compute the similarity score and the gain with

Regularization term $\lambda = 0$

Height (m)	Favorite Color	Gender	PR0
1.6	Blue	Male	16.8
1.6	Green	Female	4.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2

16.8, 4.8, -15.2, 1.8, 5.8, -14.2

$$SS = \frac{(16.8+4.8-15.2+1.8+5.8-14.2)^2}{6+\lambda} = 0.0067$$

Gender=F

4.8, -15.2, -14.2

16.8, 1.8, 5.8

$$SS = \frac{(4.8-15.2-14.2)^2}{3+\lambda} = 201.72 \qquad SS = \frac{(16.8+1.8+5.8)^2}{3+\lambda} = 198.45$$

$$\text{Gain} = 201.72 + 198.45 - 0.0067 = 400.16$$

Height<1.6

-15.2, 5.8, -14.2

16.8, 4.8, 1.8

$$\text{Gain} = 185.65 + 182.52 - 0.0067 = 368.16$$

Not Blue

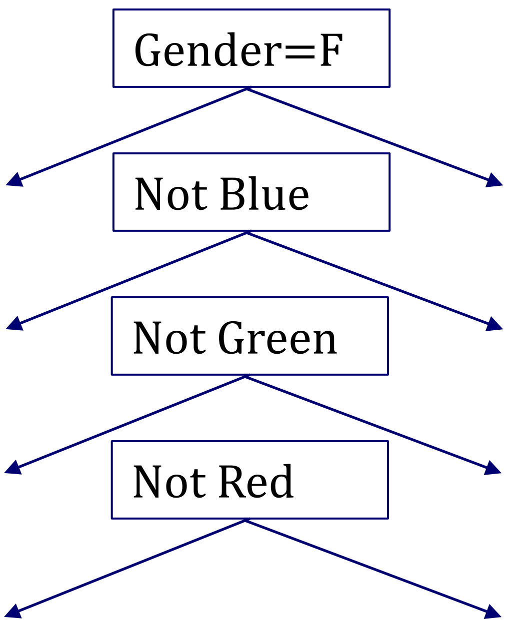
4.8, 1.8, 5.8

16.8, -15.2, -14.2

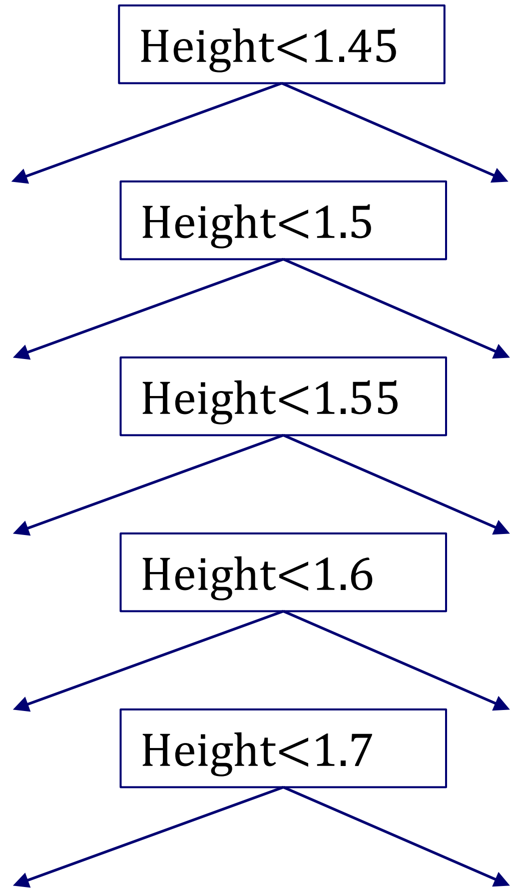
$$\text{Gain} = 51.25 + 52.92 - 0.0067 = 104.16$$

XGBoost Tree

Height (m)	Favorite Color	Gender	PR0
1.6	Blue	Male	16.8
1.6	Green	Female	4.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2



If numerical, sort and take average between adjacent

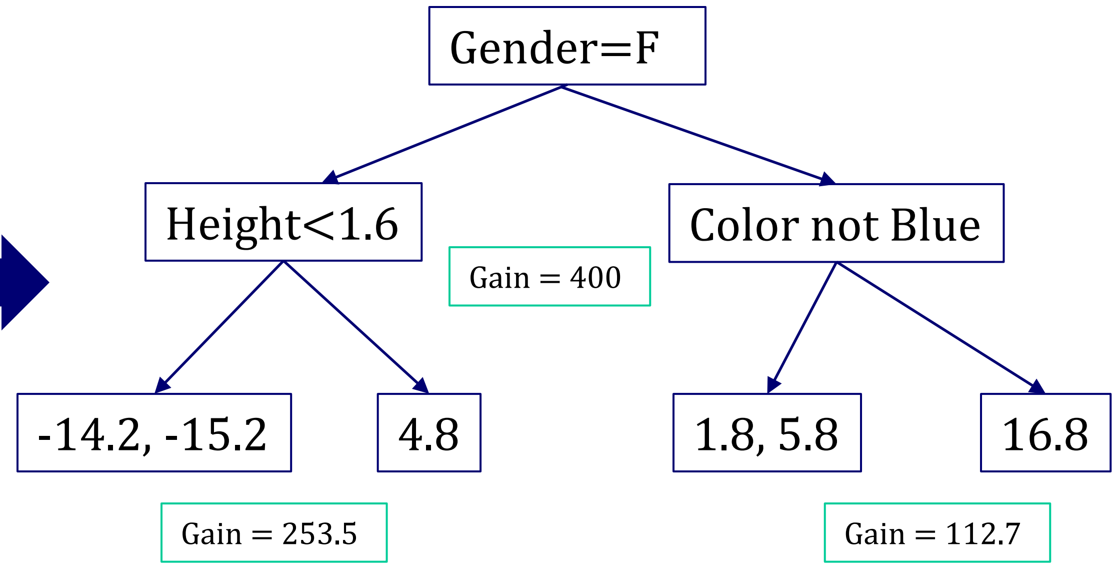


These can be re-used in building the DT.

XGBoost Tree

Fit PR0 into an XGBoost tree (up to four leaves)

Height (m)	Favorite Color	Gender	PR0
1.6	Blue	Male	16.8
1.6	Green	Female	4.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2



1) Pruning
2) Regularization

Pruning Term γ

If the gain $< \boldsymbol{\gamma}$, then we prune the tree!

The tree could be pruned even if we set $\boldsymbol{\gamma = 0}$.

Regularization Term λ

If the regularization term grows large, then the gain will be decreased, and thus the node becomes harder to split and easier to prune. This is used to prevent overfitting (of the training data).

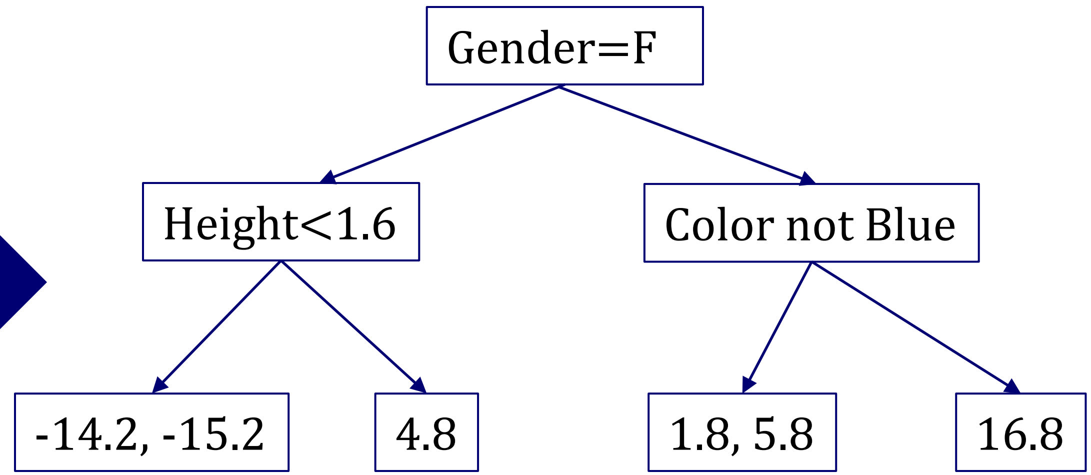
The regularization affects children nodes more than their parent node.

XGBoost Tree

Fit PR0 into an XGBoost tree (up to four leaves)

$$\lambda = 0, \gamma = 0$$

Height (m)	Favorite Color	Gender	PR0
1.6	Blue	Male	16.8
1.6	Green	Female	4.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2



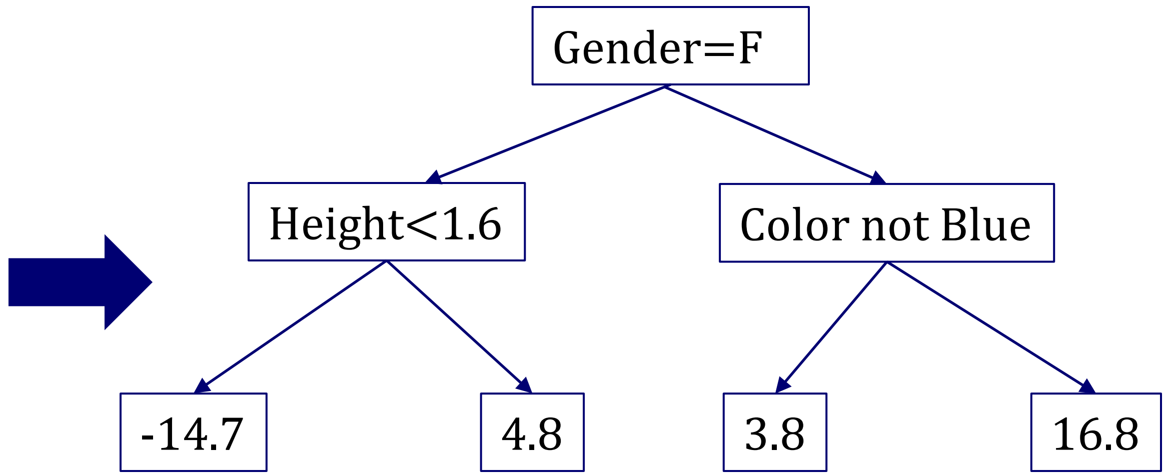
Pseudo Residual (PR) = True Value – Predicted Value

XGBoost Output

Output Value = Sum of PR / (#PRs + regularization term)

$$\boxed{\lambda = 0, \gamma = 0}$$

Height (m)	Favorite Color	Gender	PR0
1.6	Blue	Male	16.8
1.6	Green	Female	4.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2



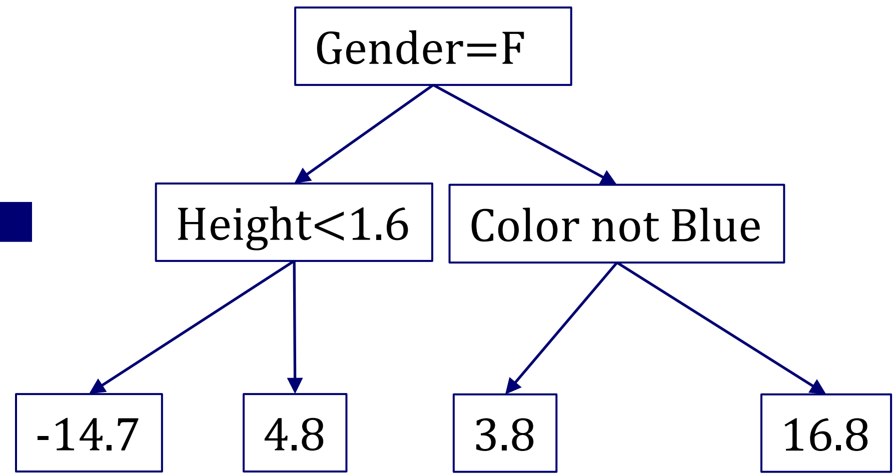
$$\frac{-14.2-15.2}{2+\lambda} = -14.7 \qquad \frac{4.8}{1+\lambda} = 4.8 \qquad \frac{1.8+5.8}{2+\lambda} = 3.8 \qquad \frac{16.8}{1+\lambda} = 16.8$$

XGBoost

Learning rate = 0.1

$$F1(x) = F0(x) + \nu_1 \times \text{Output of DT}(x)$$

Height (m)	Favorite Color	Gender	Weight (kg)	F0
1.6	Blue	Male	88	71.2
1.6	Green	Female	76	71.2
1.5	Blue	Female	56	71.2
1.8	Red	Male	73	71.2
1.5	Green	Male	77	71.2
1.4	Blue	Female	57	71.2



$$F1((1.6, \text{Blue}, \text{Male})) = 71.2 + 0.1 \times 16.8 = 72.9$$
$$F1((1.6, \text{Green}, \text{Female})) = 71.2 + 0.1 \times 4.8 = 71.7$$
$$F1((1.5, \text{Blue}, \text{Female})) = 71.2 + 0.1 \times -14.7 = 69.7$$
$$F1((1.8, \text{Red}, \text{Male})) = 71.2 + 0.1 \times 3.8 = 71.6$$
$$F1((1.5, \text{Green}, \text{Male})) = 71.2 + 0.1 \times 3.8 = 71.6$$
$$F1((1.4, \text{Blue}, \text{Female})) = 71.2 + 0.1 \times -14.7 = 69.7$$

XGBoost

So after building the first DT, we obtain…

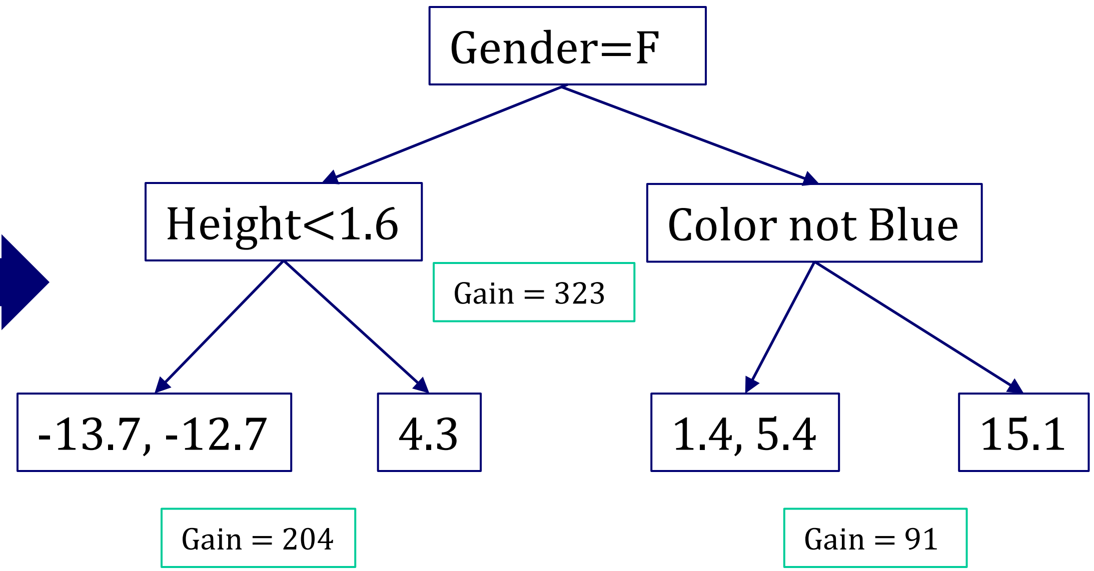
Height (m)	Favorite Color	Gender	Weight (kg)	F0	PR0	F1	PR1
1.6	Blue	Male	88	71.2	16.8	72.9	15.1
1.6	Green	Female	76	71.2	4.8	71.7	4.3
1.5	Blue	Female	56	71.2	-15.2	69.7	-13.7
1.8	Red	Male	73	71.2	1.8	71.6	1.4
1.5	Green	Male	77	71.2	5.8	71.6	5.4
1.4	Blue	Female	57	71.2	-14.2	69.7	-12.7

XGBoost

Fit PR1 into an XGBoost tree (up to four leaves)

$$\boxed{\lambda = 0, \gamma = 0}$$

Height (m)	Favorite Color	Gender	PR1
1.6	Blue	Male	15.1
1.6	Green	Female	4.3
1.5	Blue	Female	-13.7
1.8	Red	Male	1.4
1.5	Green	Male	5.4
1.4	Blue	Female	-12.7

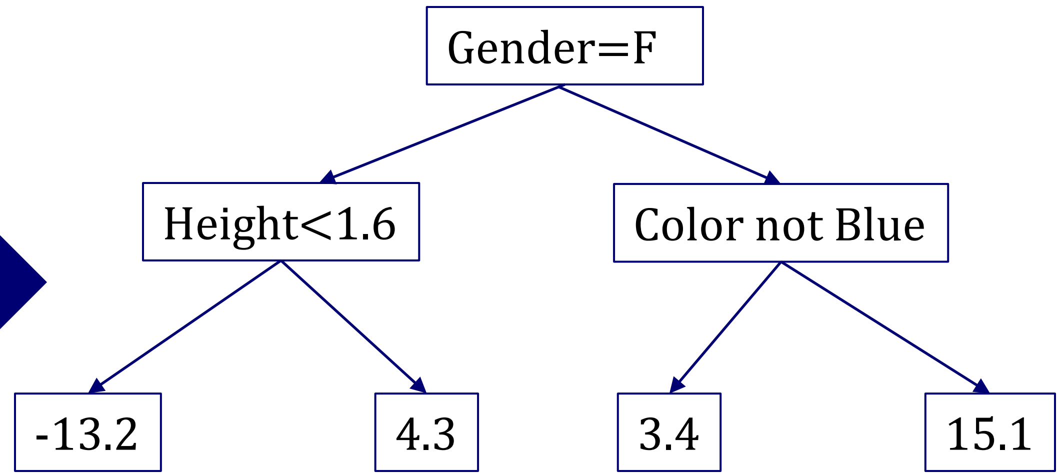


XGBoost

Output Value = Sum of PR / (#PRs + regularization term)

$\boxed{\lambda = 0, \gamma = 0}$

Height (m)	Favorite Color	Gender	PR1
1.6	Blue	Male	15.1
1.6	Green	Female	4.3
1.5	Blue	Female	-13.7
1.8	Red	Male	1.4
1.5	Green	Male	5.4
1.4	Blue	Female	-12.7

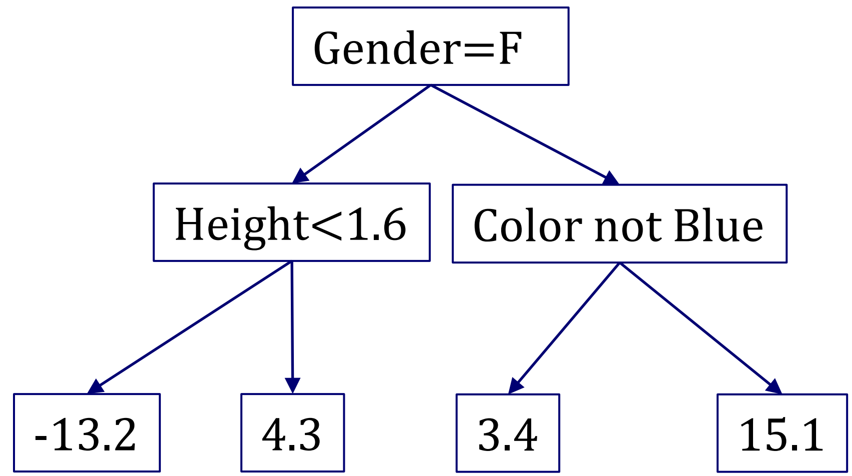


XGBoost

Learning rate = 0.1

$$F2(x) = F1(x) + \nu_2 \times \text{Output of DT}(x)$$

Height (m)	Favorite Color	Gender	Weight (kg)	F1
1.6	Blue	Male	88	72.9
1.6	Green	Female	76	71.7
1.5	Blue	Female	56	69.7
1.8	Red	Male	73	71.6
1.5	Green	Male	77	71.6
1.4	Blue	Female	57	69.7



$$F2((1.6, \text{Blue}, \text{Male})) = 72.9 + 0.1 \times 15.1 = 74.4$$
$$F2((1.6, \text{Green}, \text{Female})) = 71.7 + 0.1 \times 4.3 = 72.1$$
$$F2((1.5, \text{Blue}, \text{Female})) = 69.7 + 0.1 \times -13.2 = 68.4$$
$$F2((1.8, \text{Red}, \text{Male})) = 71.6 + 0.1 \times 3.4 = 71.9$$
$$F2((1.5, \text{Green}, \text{Male})) = 71.6 + 0.1 \times 3.4 = 71.9$$
$$F2((1.4, \text{Blue}, \text{Female})) = 69.7 + 0.1 \times -13.2 = 68.4$$

XGBoost

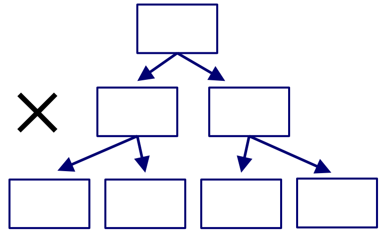
So after building the second DT, we obtain…

Height (m)	Favorite Color	Gender	Weight (kg)	F0	PR0	F1	PR1	F2	PR2
1.6	Blue	Male	88	71.2	16.8	72.9	15.1	74.4	13.6
1.6	Green	Female	76	71.2	4.8	71.7	4.3	72.1	3.9
1.5	Blue	Female	56	71.2	-15.2	69.7	-13.7	68.4	-12.4
1.8	Red	Male	73	71.2	1.8	71.6	1.4	71.9	1.1
1.5	Green	Male	77	71.2	5.8	71.6	5.4	71.9	5.1
1.4	Blue	Female	57	71.2	-14.2	69.7	-12.7	68.4	-11.4

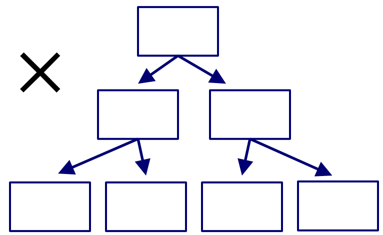
Notice the PR’s are shrinking: Small steps towards the right direction!

XGBoost

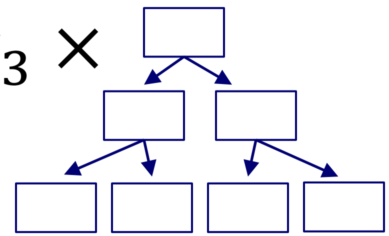
$$Fm = F0 + \nu_1 \times$$



$$+ \ \nu_2 \times$$



$$+ \ \nu_3 \times$$



$$+ \ \ldots$$

Fit the new PR into XGBoost Tree (with regularization and pruning)

Stop until the pre-specified #DTs or the PR stops improving!

Python Time!

- pip install xgboost
- Import xgboost as xgb
- Install graphviz-2.38.zip for visualization

XBoost

- Works exceptionally well in practice
- Won a series of Kaggle competitions
- More robust and explainable than DNN
- Comes with regularization and pruning