LEC004 Demand Forecasting

VG441 SS2020

Cong Shi Industrial & Operations Engineering University of Michigan

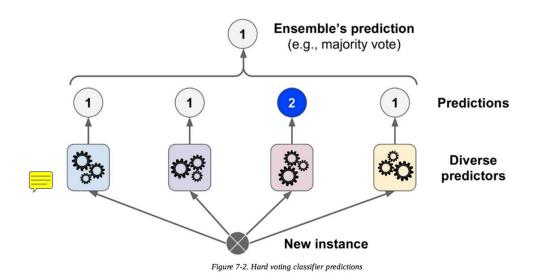
Ensemble Learning

"The wisdom of the crowd is the collective opinion of a group of individuals rather than that of a single expert."

"A group of predictors is called an ensemble. Therefore this Machine Learning technique is known as Ensemble Learning. Voilá!"

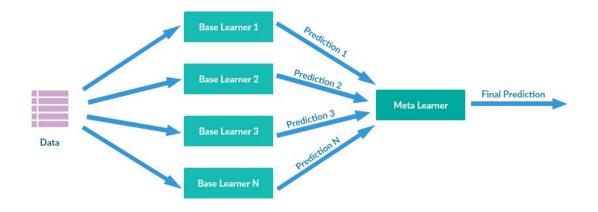
"Ensemble methods work best when the predictors are as independent of one another as possible. One way to get diverse classifiers is to train them using very different algorithms. This increases the chance that they will make very different types of errors, improving the ensemble's accuracy."

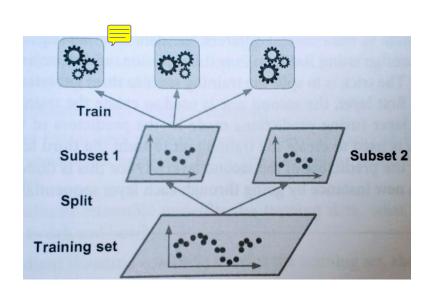
Hard voting classifier (for classification)

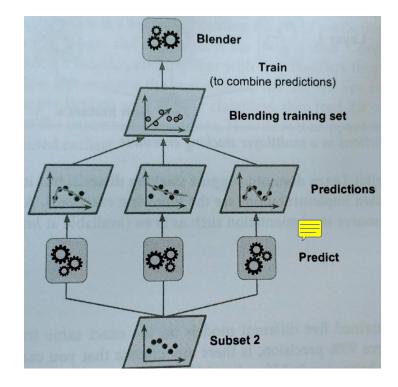


Averaging or weighted averaged (for regression)

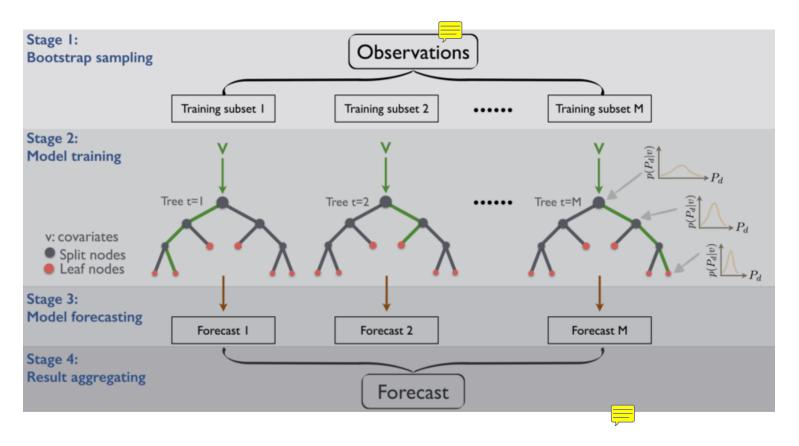
Stacking





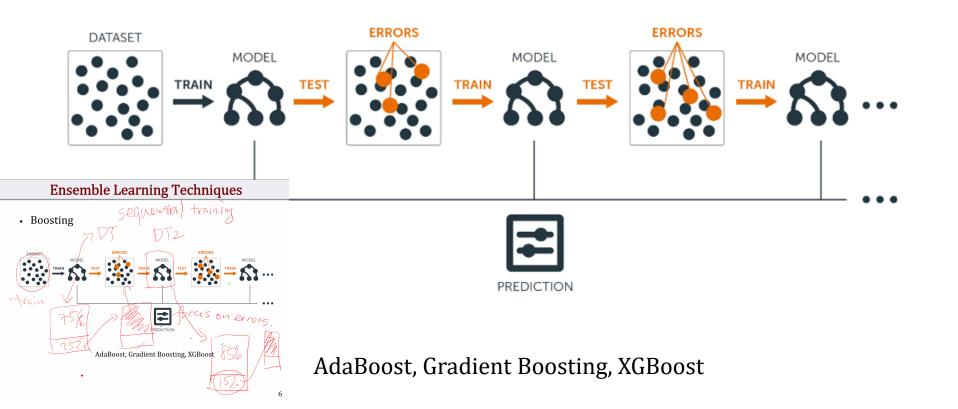


Bagging =



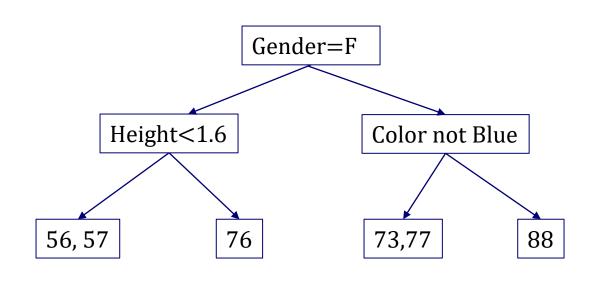
Random Forest

Boosting



Decision Tree

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



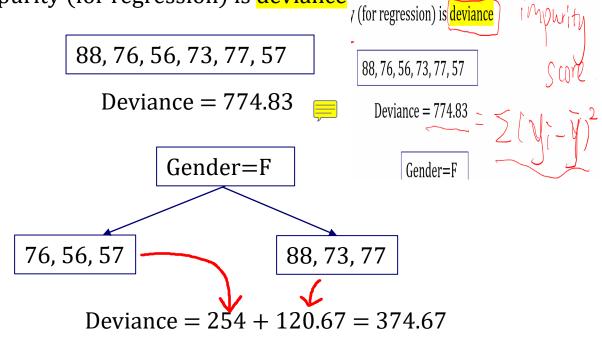
Question 1: How do we determine the next node (starting from root)?

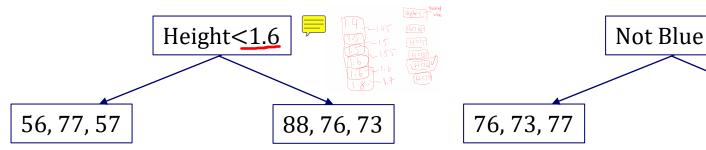
Question 2: Should we split at the current node?

How to determine and split a node?

Measure of impurity (for regression) is deviance

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





Deviance = 280.67 + 126 = 406.67

Deviance = 8.67 + 662 = 670.67

88, 56, 57

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



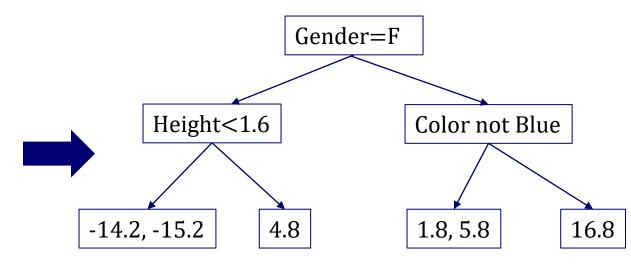
Height (m)	Favorite Color	Gender	Weight (kg)	F0	PR0
1.6	Blue	Male	88	71.2	16.8
1.6	Green	Female	76	71.2	4.8
1.5	Blue	Female	56	71.2	-15.2
1.8	Red	Male	73	71.2	1.8
1.5	Green	Male	77	71.2	5.8
1.4	Blue	Female	57	71.2	-14.2

Pseudo Residual (PR) = True Value - Predicted Value



Fit PR0 into a decision tree (up to four leaves)

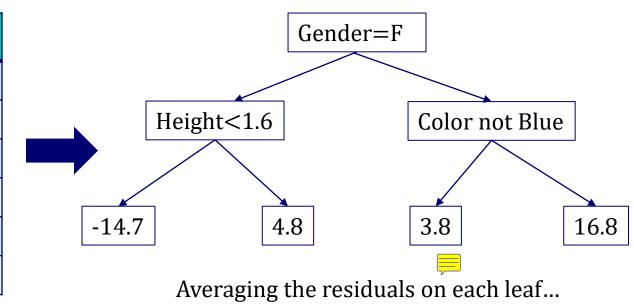
Height (m)	Favorite Color	Gender	PR0
1.6	Blue	Male	16.8
1.6	Green	Female	4.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2



Pseudo Residual (PR) = True Value - Predicted Value

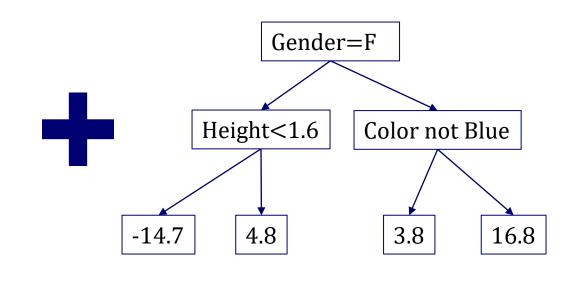
Fit PR0 into a decision tree (up to four leaves)

Height (m)	Favorite Gender Color		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



F1(x) = F0(x) + γ_1 × 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, Blue, Male)) =
$$71.2 + 0.1 \times 16.8 = 72.9$$

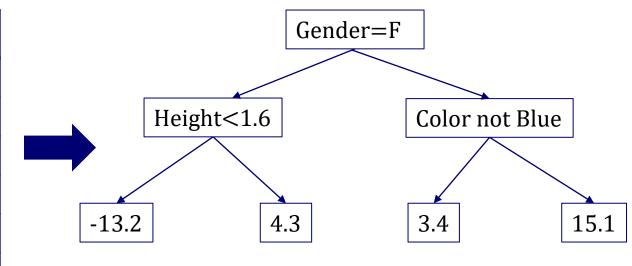
F1((1.6, Green, Female)) = $71.2 + 0.1 \times 4.8 = 71.7$
F1((1.5, Blue, Female)) = $71.2 + 0.1 \times -14.7 = 69.7$
F1((1.8, Red, Male)) = $71.2 + 0.1 \times 3.8 = 71.6$
F1((1.5, Green, Male)) = $71.2 + 0.1 \times 3.8 = 71.6$
F1((1.4, Blue, Female)) = $71.2 + 0.1 \times -14.7 = 69.7$

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

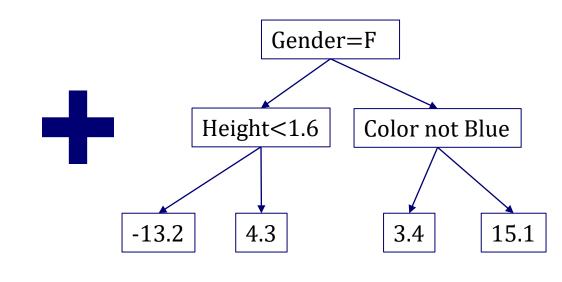
Fit PR1 into a decision tree (up to four leaves)

Height (m)	Favorite Gender Color		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



F2(x) = F1(x) + γ_2 × 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



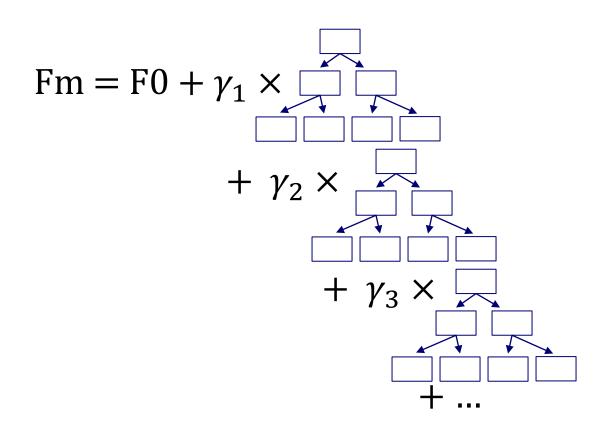
F2((1.6, Blue, Male)) =
$$72.9 + 0.1 \times 15.1 = 74.4$$

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

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!



Fit the new PR into DT

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

Python Time!

from sklearn import ensemble



Some Mathematics...

Input: training set $\{(x_i, y_i)\}_{i=1}^n$, a differentiable loss function L(y, F(x)), number of iterations M.

Algorithm:

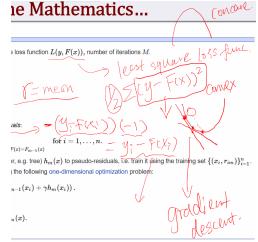


1. Initialize model with a constant value:

$$F_0(x) = rg \min_{\gamma} \sum_{i=1}^n L(y_i, \underline{\gamma}).$$

- 2. For m = 1 to M:
 - 1. Compute so-called pseudo-residuals:

$$r_{im} = -iggl[rac{\partial L(y_i,F(x_i))}{\partial F(x_i)}iggr]_{F(x)=F_{m-1}(x)} \quad ext{for } i=1,\ldots,n.$$



- 2. Fit a base learner (or weak learner, e.g. tree) $h_m(x)$ to pseudo-residuals, i.e. train it using the training set $\{(x_i, r_{im})\}_{i=1}^n$.
- 3. Compute multiplier γ_m by solving the following one-dimensional optimization problem:

$$\gamma_m = rg \min_{\gamma} \sum_{i=1}^n L\left(y_i, \underbrace{F_{m-1}(x_i) + \gamma h_m(x_i)}
ight).$$

4. Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

3. Output $F_M(x)$.

Works exceptionally well in practice



Won a series of Kaggle competitions

More robust and explainable than DNN