

For Regression:

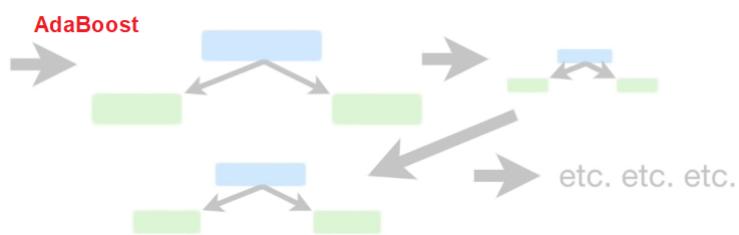
Gradient Boost和AdaBoost的区别:

AdaBoost:通过引入上一个stump的误差建立下一个stump, 当stump的数量达到预设值或perfect fit时停止

GradientBoost:

用叶节点作为initial guess

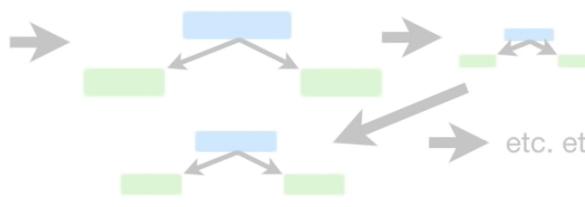
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc...	etc...	etc...	etc...



Gradient Boost

This leaf represents an initial guess for the **Weights** of all of the samples.

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc...	etc...	etc...	etc...

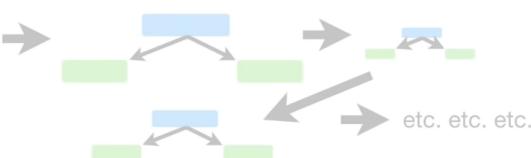


73.3

When trying to **Predict** a continuous value like **Weight**, the first guess is the average value.

initial guess的值是所有样本的平均值。

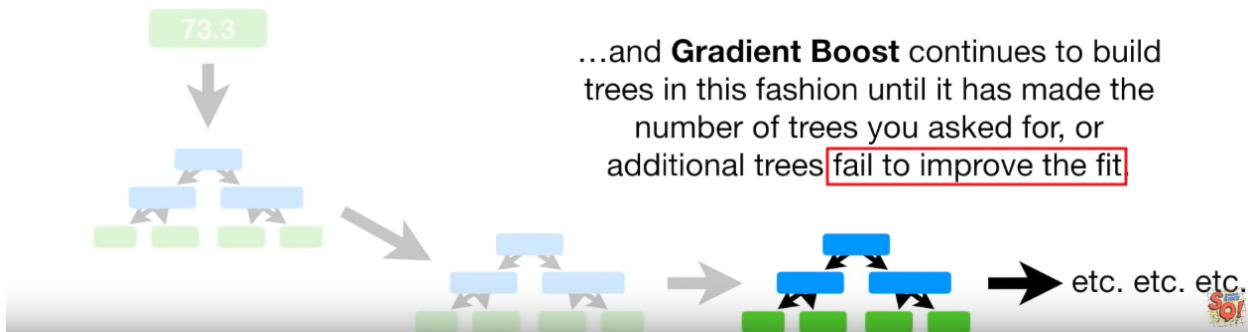
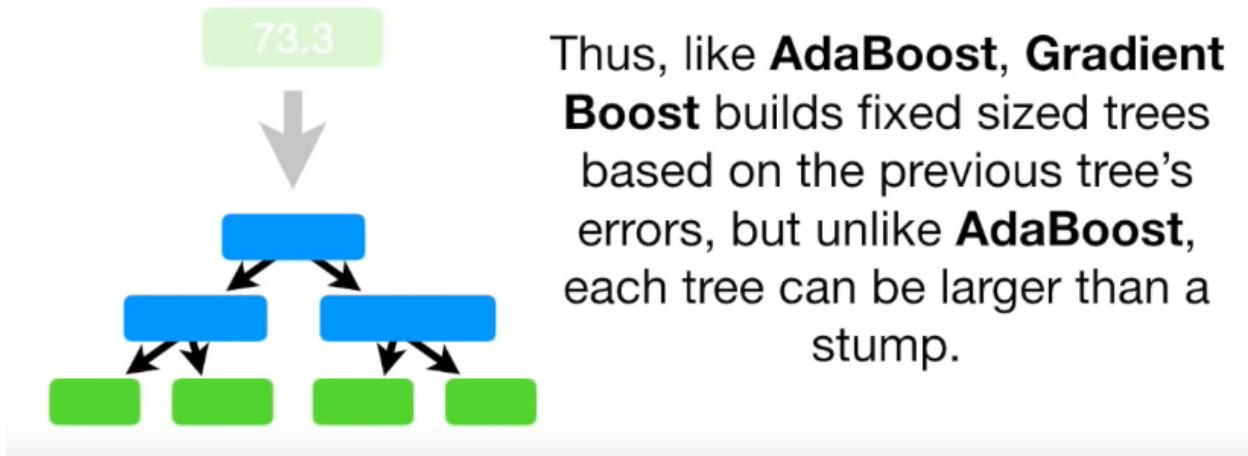
Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc...	etc...	etc...	etc...



73.3

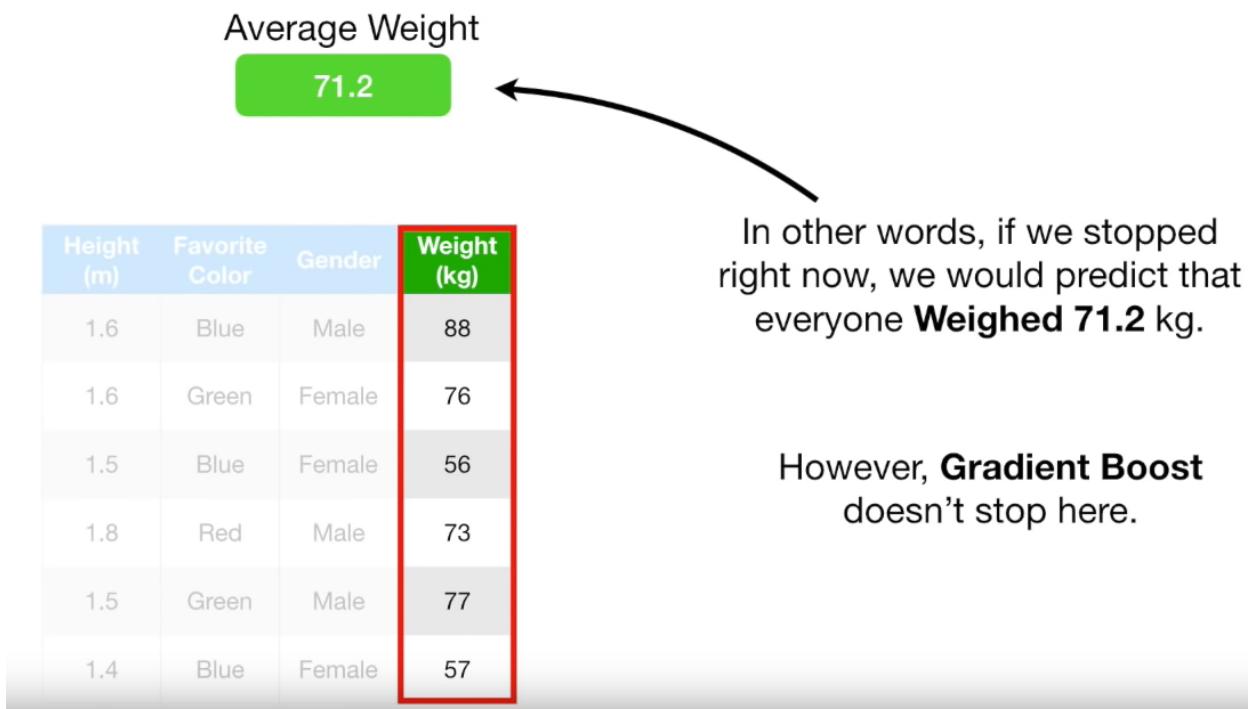
Like **AdaBoost**, this tree is based on the errors made by the previous tree...

和AdaBoost类似，后面的树也是根据前面的树的误差构造的。和AdaBoost不同，Gradient Boost中的树更大，一般有4、8、32个叶节点（每个树都有相同的叶节点数）。

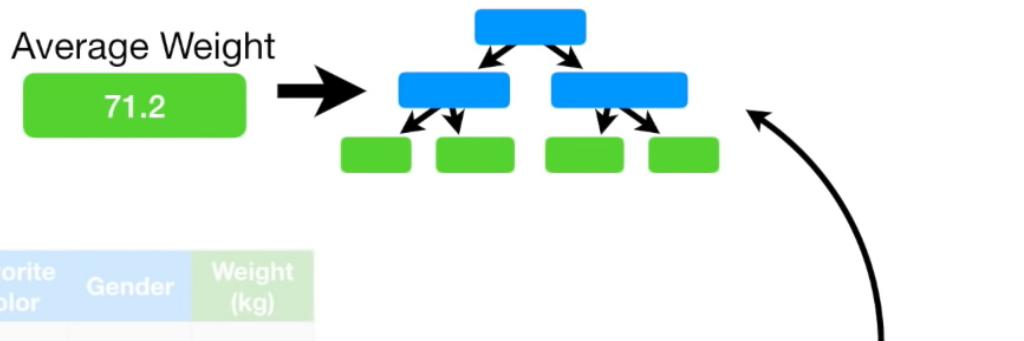


Gradient Boost预测weight的求解过程：

(1) 把training set中所有样本的weight的平均值作为initial guess



(2) 求initial guess和真实数据的误差Error



The next thing we do is build a tree based on the errors from the first 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



The errors that the previous tree made are the differences between the **Observed Weights** and the **Predicted Weight, 71.2**.

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

(Observed Weight - Predicted Weight)



Average Weight

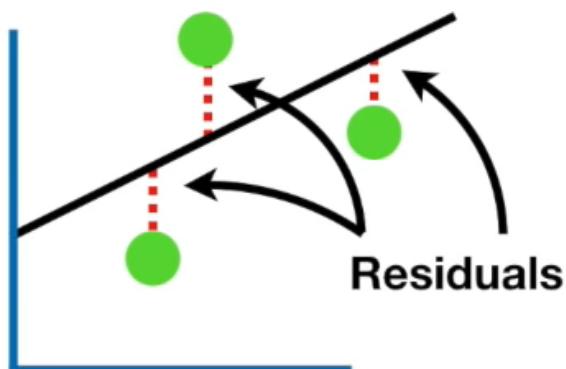
71.2

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
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	

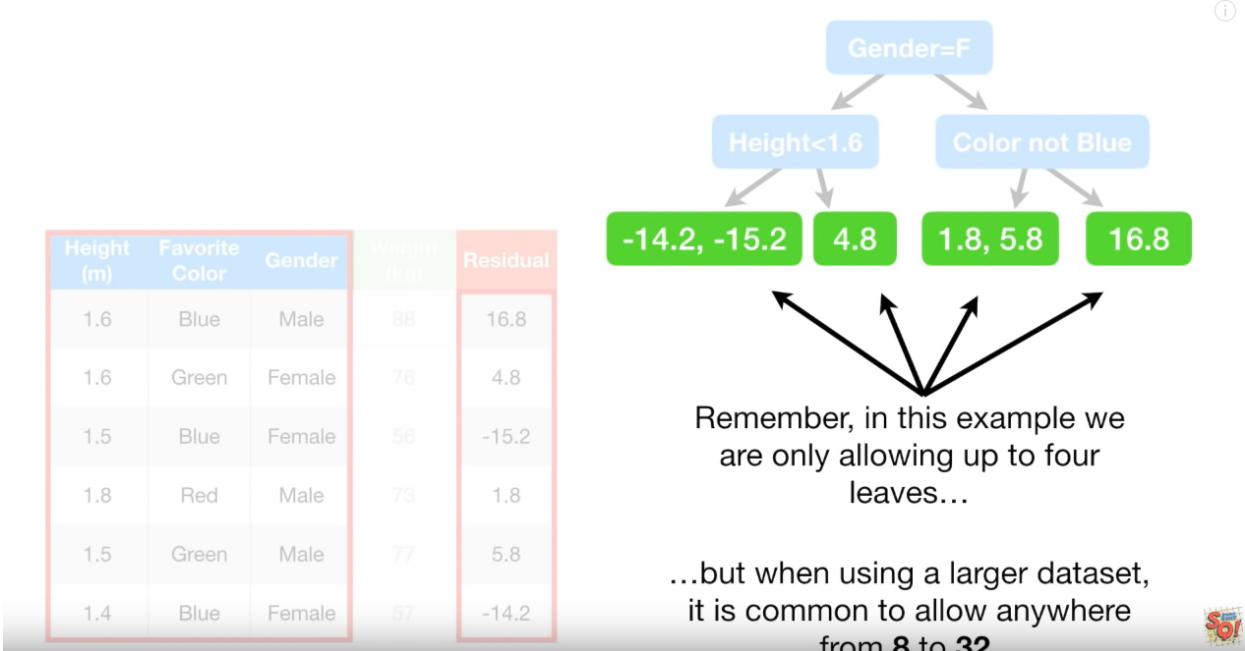
...and save the difference, which is called a **Pseudo Residual**, in a new column.

$$(88 - 71.2) = 16.8$$

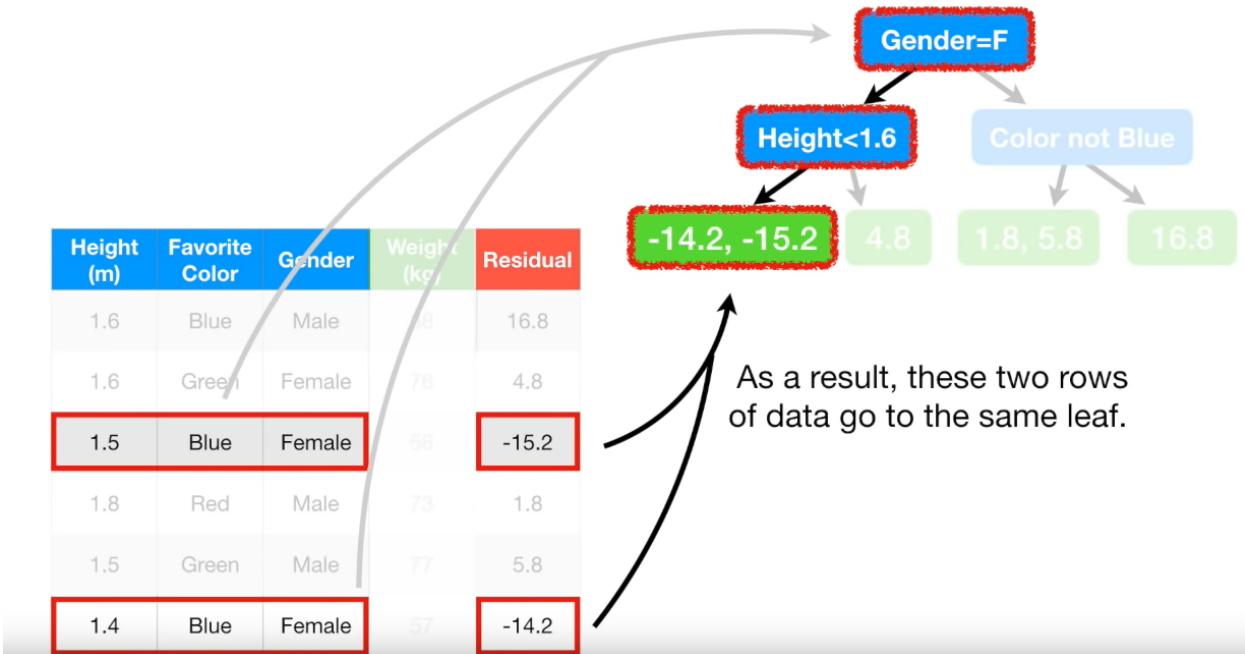
**NOTE:** The term **Pseudo Residual** is based on **Linear Regression**, where the difference between the **Observed** values and the **Predicted** values results in **Residuals**.



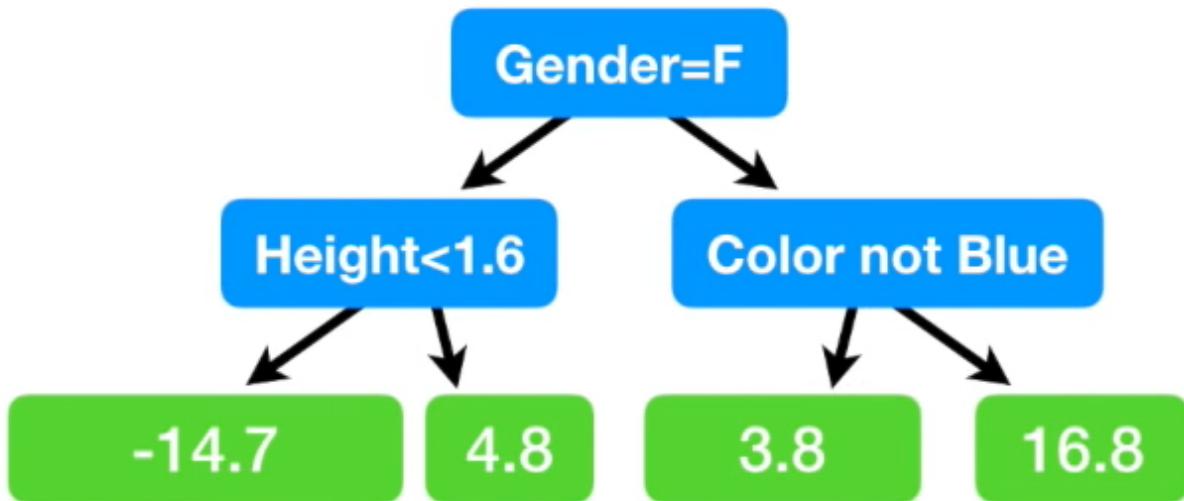
(2) 使用decision tree对residual进行分类



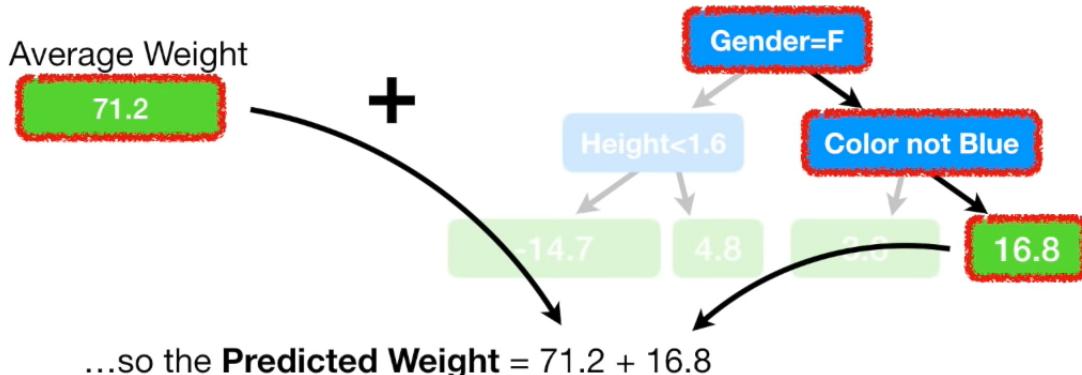
因为限定了叶节点数为4，因此，叶节点数比residual的数量要少。



将同一个叶节点的residual用他们的均值代替



(3) 使用tree进行预测



Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88



Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

...which is the same as the **Observed Weight**.

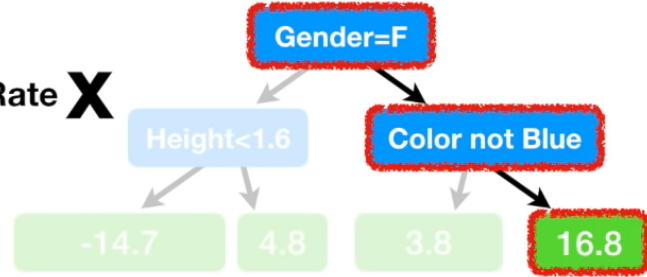
这样会造成over fit的问题（低偏差，高方差）。

解决方法是加上一个学习率：

Average Weight

71.2

+ Learning Rate  $\times$



Gradient Boost deals with this problem by using a **Learning Rate** to scale the contribution from the new tree.

The **Learning Rate** is a value between **0** and **1**.

假设学习率为0.1：

Average Weight

71.2

0.1  $\times$

$$\text{Predicted Weight} = 71.2 + (0.1 \times 16.8) = 72.9$$

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88

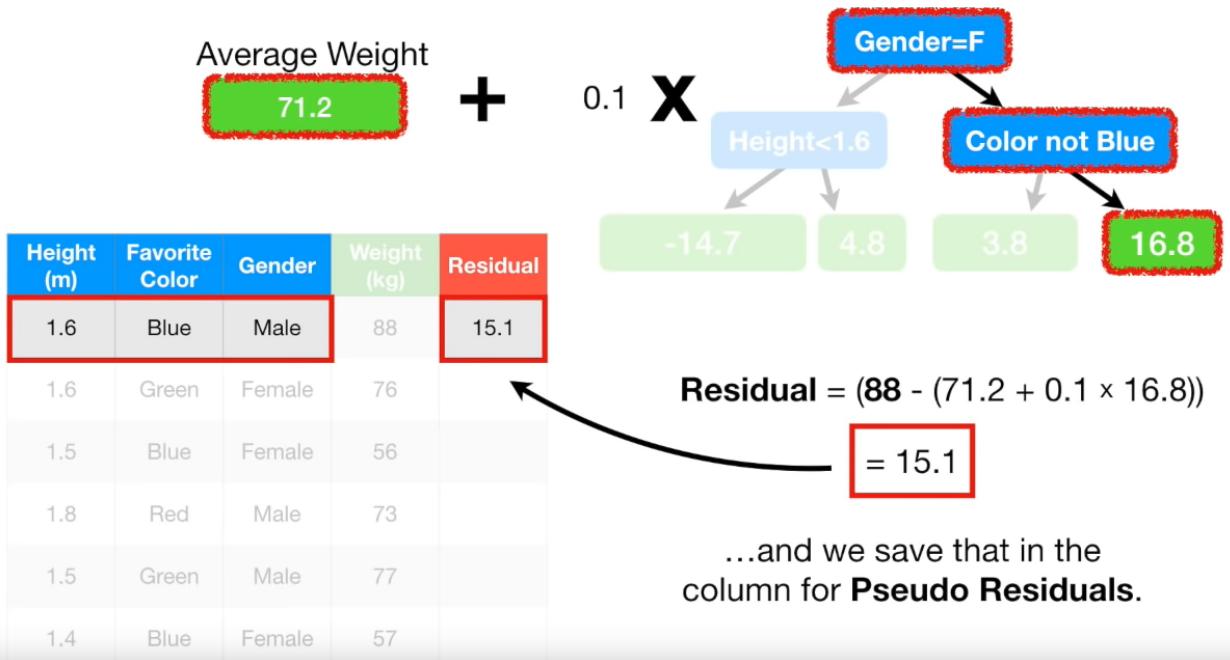
...but it's a little bit better than the **Prediction** made with just the original leaf, which predicted that all samples would weigh **71.2**.

虽然72.9没有88那么精确，但是比initial guess的71.2要好。

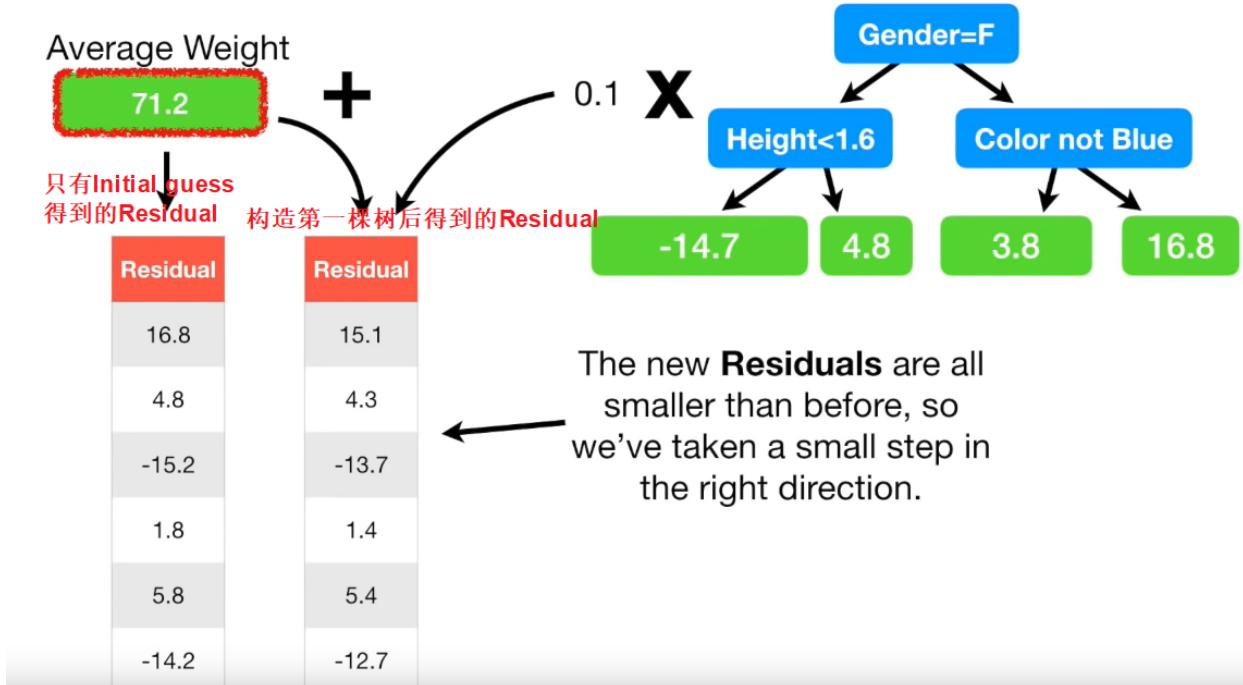
According to the dude that invented **Gradient Boost**, Jerome Friedman, empirical evidence shows that taking lots of small steps in the right direction results in better **Predictions** with a **Testing Dataset**, i.e. lower **Variance**.

Gradient Boost的目的就是往正确的方向，一步一步逼近。

### (3) 根据上一个tree构造下一个tree

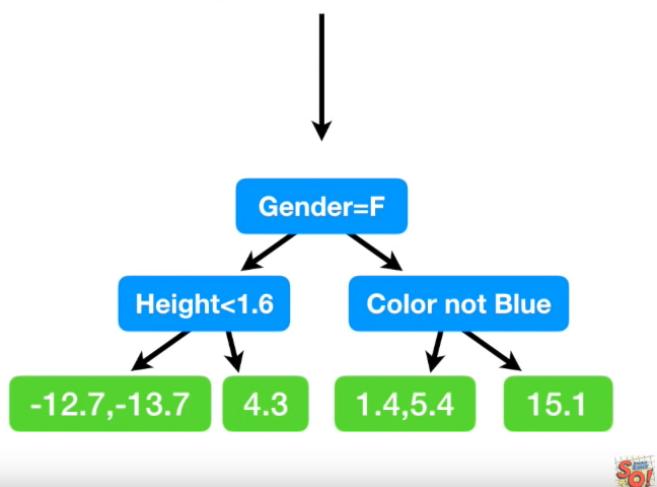


通过上一个tree得出每个sample的预测值，再得到目前对应的residual。

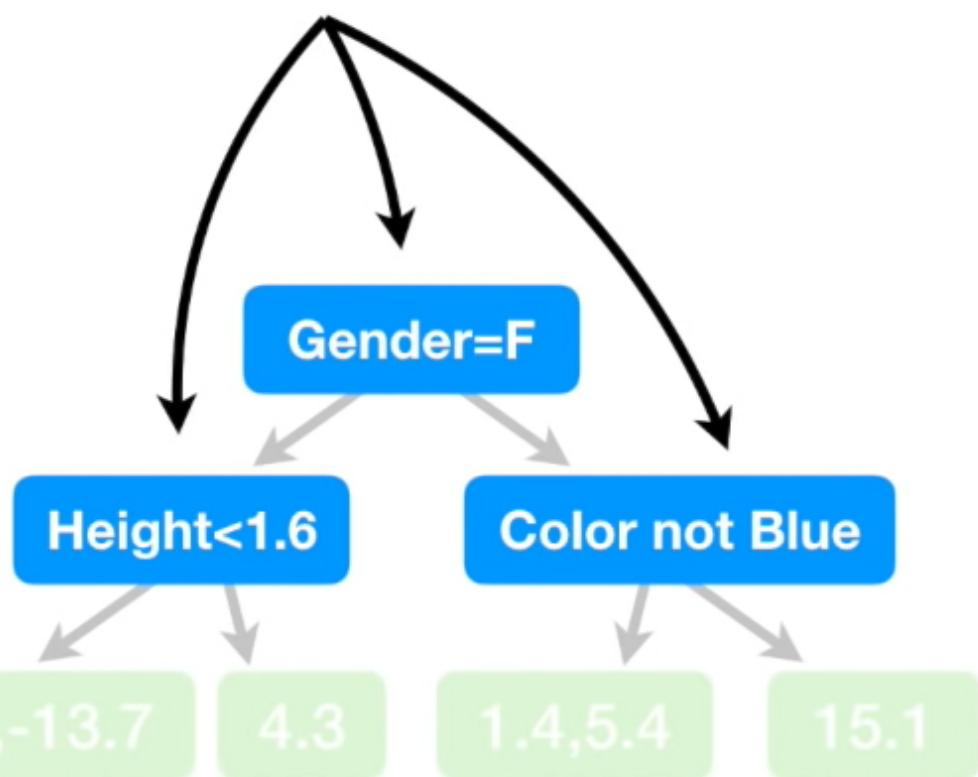


And here's the new tree!

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	15.1
1.6	Green	Female	76	4.3
1.5	Blue	Female	96	-13.7
1.8	Red	Male	73	1.4
1.5	Green	Male	77	5.4
1.4	Blue	Female	67	-12.7



**NOTE:** In this simple example the branches are the same as before. However, in practice, the trees can be different each time.



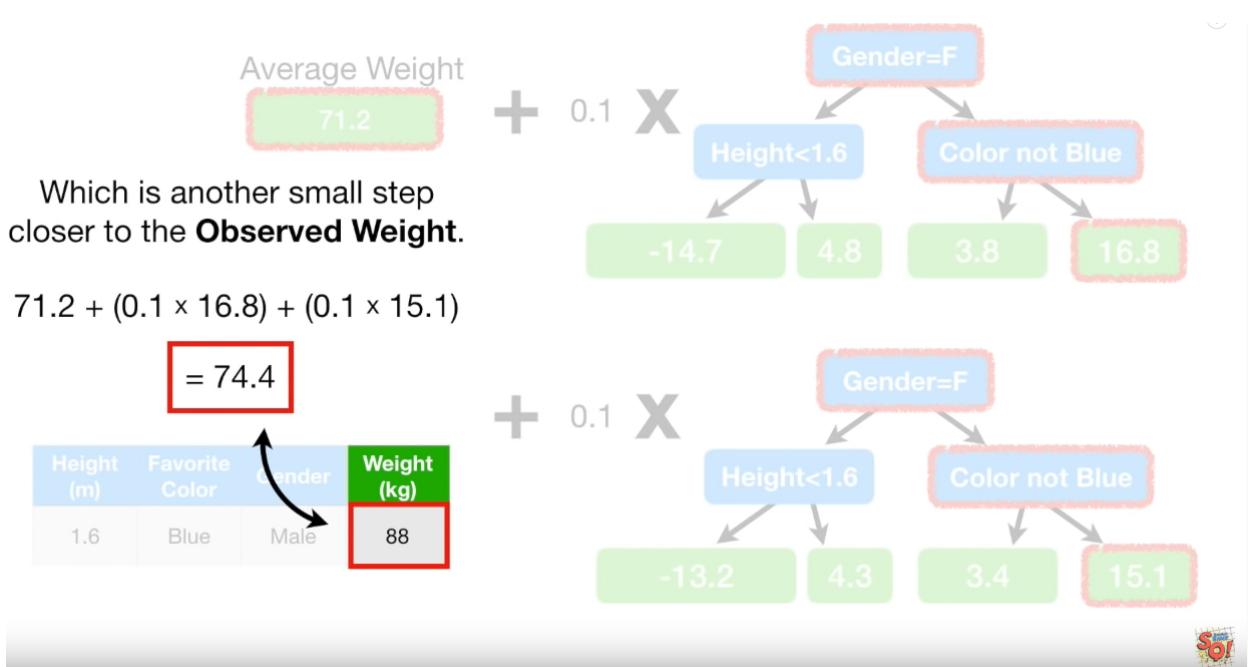
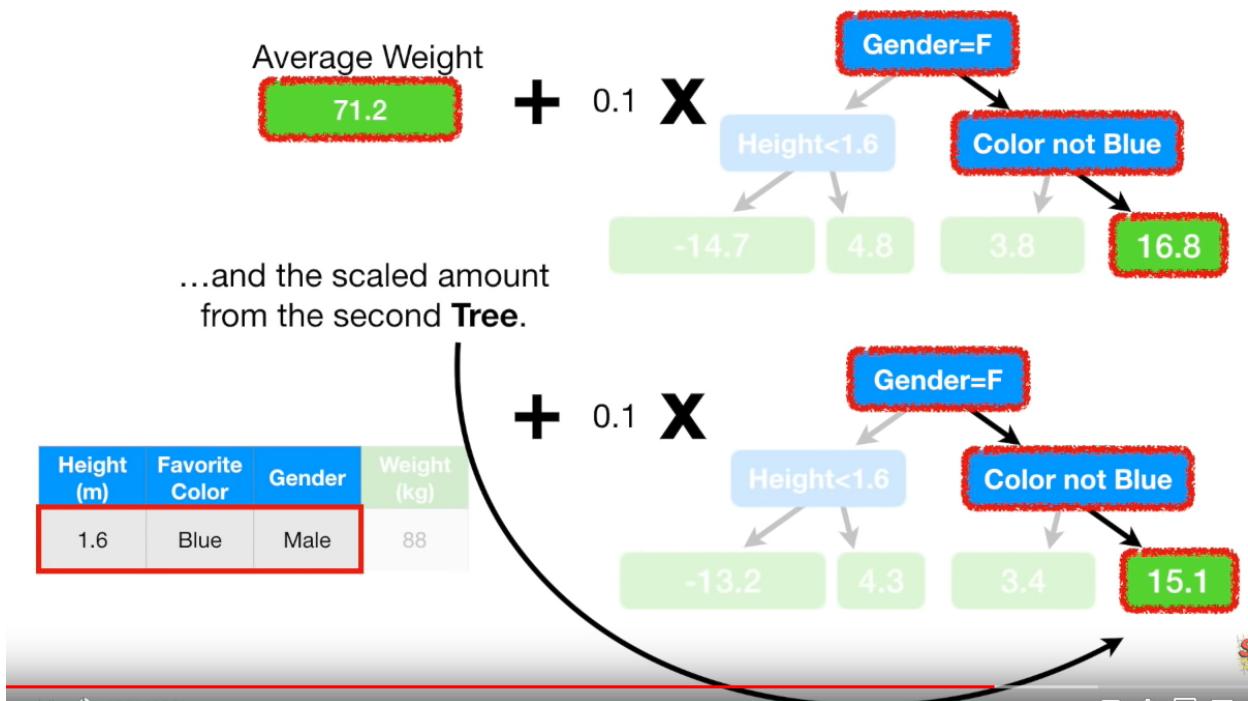
分支条件跟上一个tree是相同的。但实际应用中，可以不同。

用均值代替：

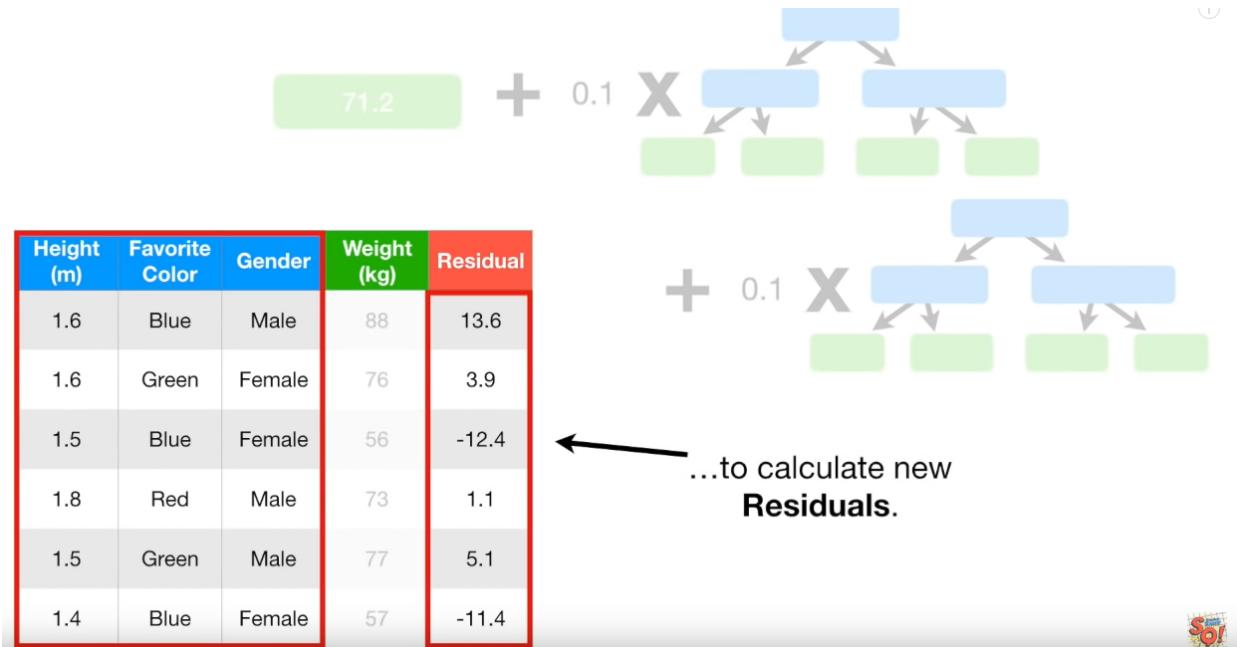
Just like before, since multiple samples ended up in these leaves, we just replace the **Residuals** with their averages.



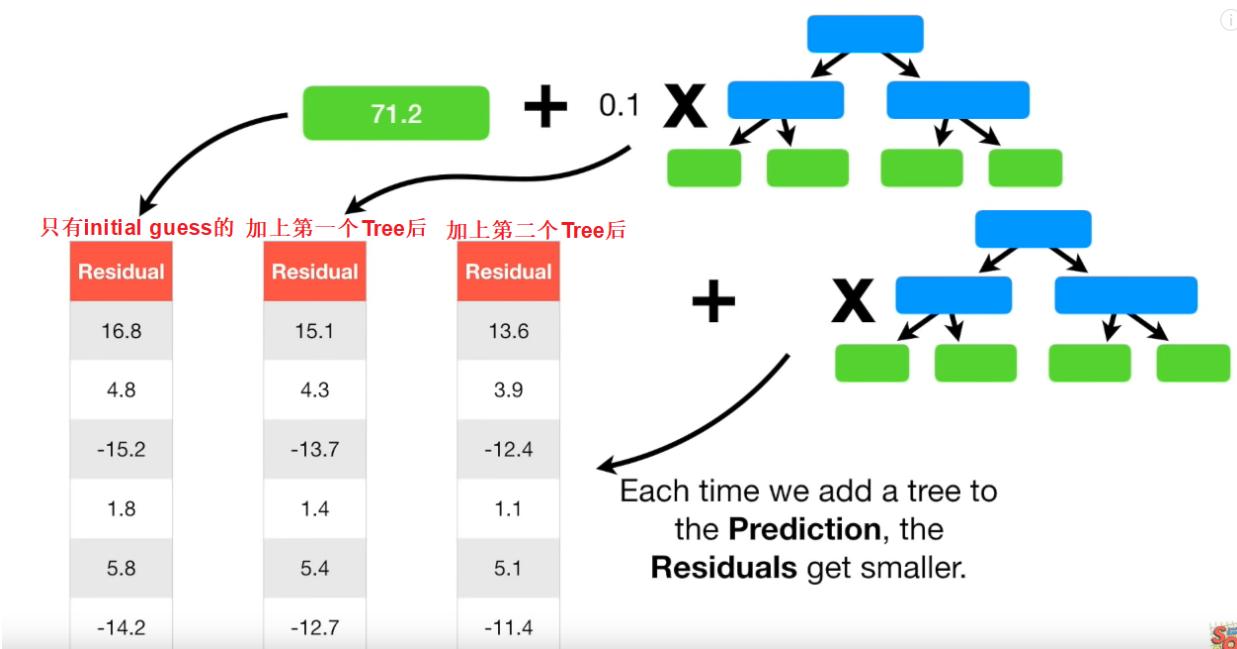
#### (4) 使用已有的两个tree进行预测



并得到最新的residual



对比一下目前为止得到的Residual:



如此类推，构造足够多的tree，直到tree的数量达到预设值或residuals没有很大的变化。

...and we keep making trees until we reach the maximum specified, or adding additional trees does not significantly reduce the size of the **Residuals**.

预测：

