



Ensembles: Gradient Boost

Ing. Juan M. Rodríguez



Gradient Boost

1. Gradient Boost crea una cadena de árboles de profundidad fija.
2. Comienza con un solo valor, un nodo hoja. Y luego calcula árboles para calcular el error cometido por el anterior.
3. Cada árbol está ponderado por un factor constante llamado *learning rate* o tasa de aprendizaje.
 - a. Los árboles están restringidos en su crecimiento



Gradient Boost: ejemplo

| 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 |

En este ejemplo, queremos determinar un valor en un rango continuo, el peso (**Weight**), por lo tanto estamos ante un problema de regresión. **Gradient Boost** puede usarse para problemas de regresión o de clasificación.

Paso 1:

Calculamos una sola hoja de un árbol, que clasifica todos los casos.

Esta hoja es el promedio.

71.2

Gradient Boost: ejemplo

| Height (m) | Favorite color | Gender | Weight (kg) | Prediction (kg) | Residual |
|------------|----------------|--------|-------------|-----------------|----------|
| 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 |

71.2

Paso 2:

Calculamos el error cometido por el estimador hasta ahora.

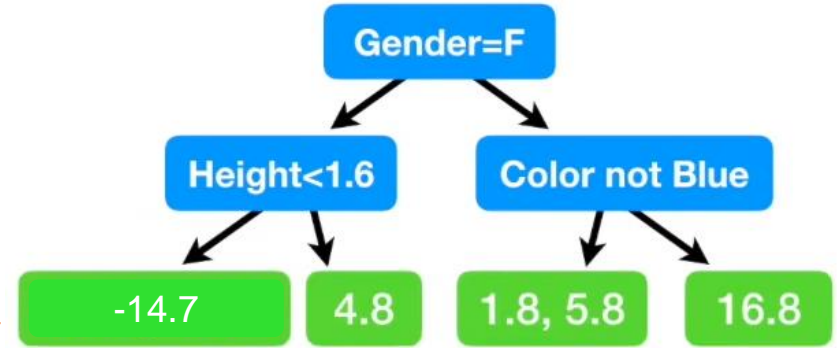
Se lo llama: **Residuo**

Residuo = (Peso observado - peso predicho)

NOTA: El término correcto sería **pseudo-residuo**. Los **residuos** son los errores en regresión lineal, acá se toma la palabra por similitud, pero se le agrega el **pseudo** para entender que es otro modelo.

Gradient Boost: ejemplo

| Height (m) | Favorite color | Gender | Residual |
|------------|----------------|--------|----------|
| 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 |



Unificamos estos dos valores
Con el promedio

$$\frac{-14.2 + (-15.2)}{2} = -14.7$$

Paso 3:

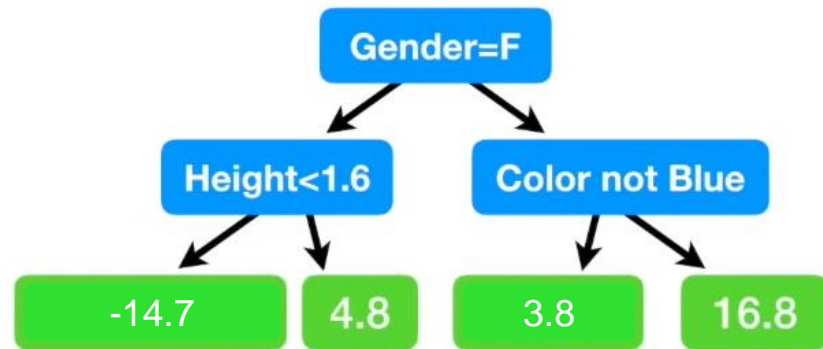
Construimos un árbol. Para los Residuos.

El árbol estará limitado a solo **4 hojas**.

En casos reales se usa un valor entre 8 y 32 hojas.

Gradient Boost: ejemplo

| Height (m) | Favorite color | Gender | Residual |
|------------|----------------|--------|----------|
| 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 |



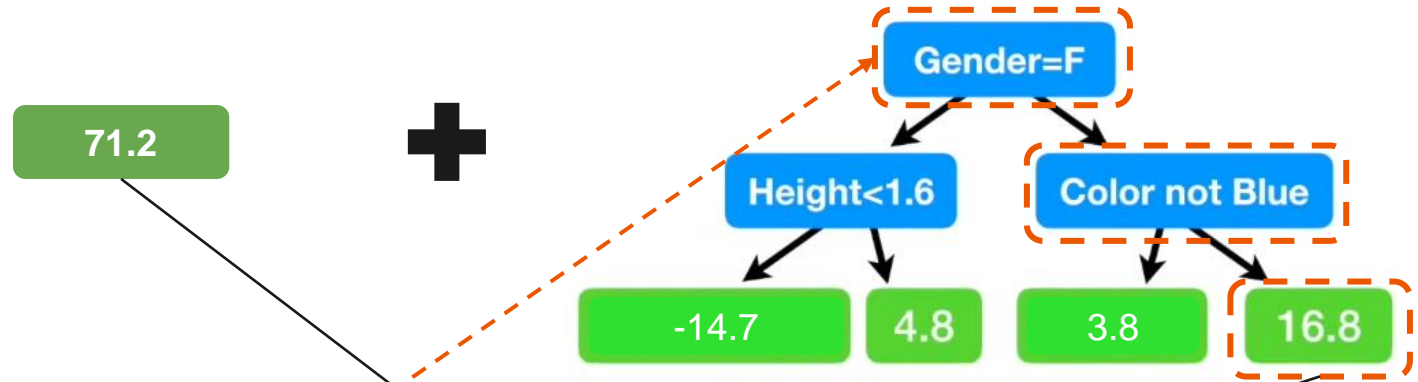
Unificamos estos
dos valores
Con el promedio

$$\frac{1.8 + 5.8}{2} = 3.8$$

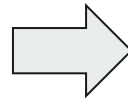
Paso 3:

Construimos un árbol. Para los Residuos.
El árbol estará limitado a solo **4 hojas**.
En casos reales se usa un valor entre 8 y 32 hojas.

Gradient Boost: ejemplo



| Height (m) | Favorite color | Gender | Weight (kg) |
|------------|----------------|--------|-------------|
| 1.6 | Blue | Male | 88 |



71.2

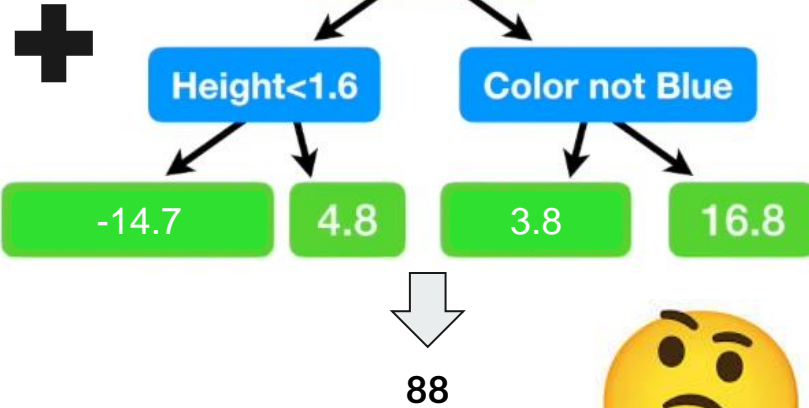
+

16.8

= 88

Gradient Boost: ejemplo

71.2



El modelo predice sospechosamente bien...

| Height (m) | Favorite color | Gender | Weight (kg) |
|------------|----------------|--------|-------------|
| 1.6 | Blue | Male | 88 |



Bias vs. Variance

Bias

El error debido al *Bias* de un modelo es simplemente la diferencia entre el valor esperado del estimador (es decir, la predicción media del modelo) y el valor real.

Cuando se dice que un modelo tiene un bias muy alto quiere decir que **el modelo es muy simple y no se ha ajustado a los datos de entrenamiento** (suele ser *underfitting*), por lo que produce un error alto en todas las muestras: entrenamiento, validación y test



Bias vs. Variance

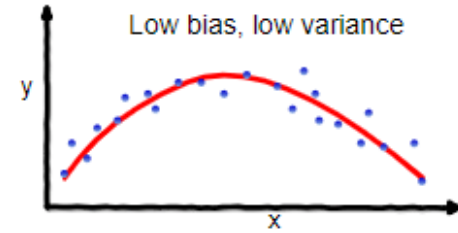
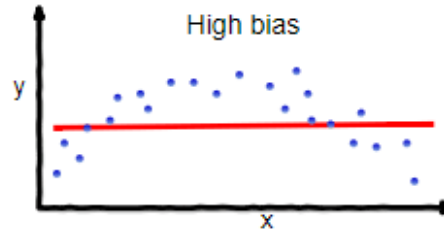
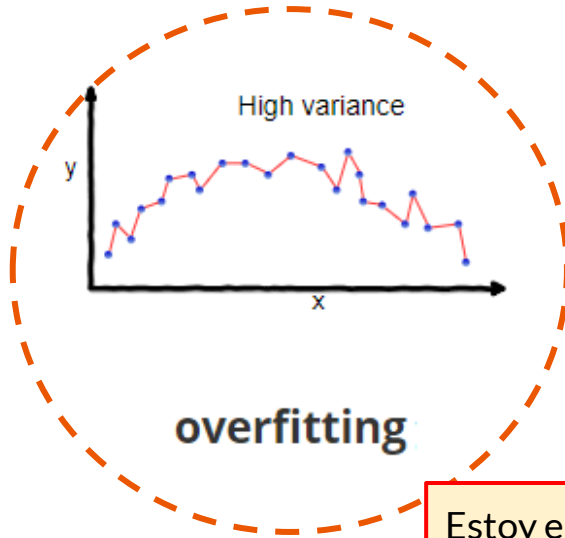
Variance (varianza)

La varianza de un estimador es cuánto varía la predicción según los datos que utilicemos para el entrenamiento.

Un modelo con **varianza baja indica que cambiar los datos de entrenamiento produce cambios pequeños en la estimación.**

Al contrario, un modelo con **varianza alta quiere decir que pequeños cambios en el dataset conlleva a grandes cambios en el output** (suele ser *overfitting*).

tradeoff bias vs variance



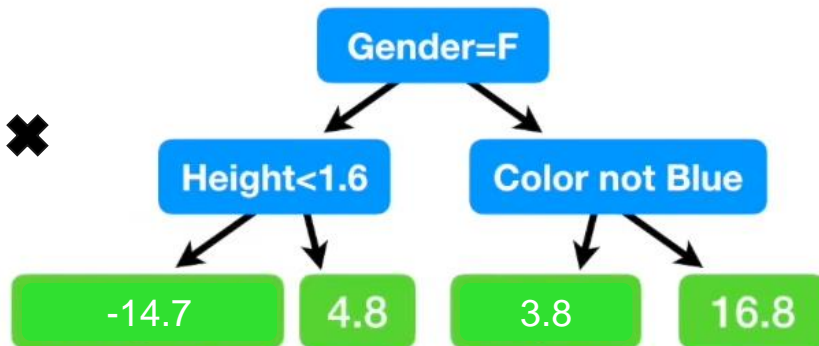
Estoy en este caso. El estimador aprendió de memoria los datos de entrenamiento pero no será bueno con nuevos datos nunca antes vistos.

Gradient Boost: ejemplo

71.2



Learning_Rate ×



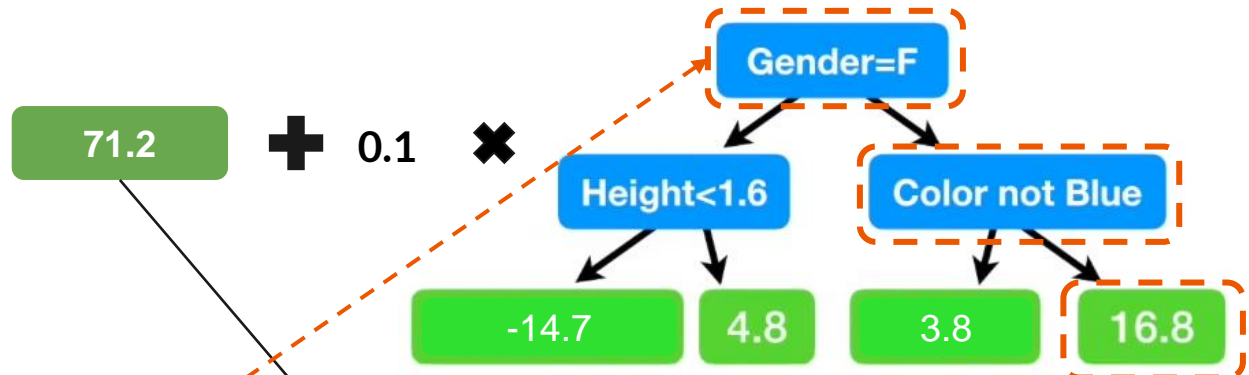
Gradient Boost maneja el **overfitting** usando un parametro llamado **Learning Rate**, para escalar la contribución de cada árbol.

Es un valor entre 0 y 1.

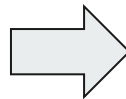
| Height (m) | Favorite color | Gender | Weight (kg) |
|------------|----------------|--------|-------------|
| 1.6 | Blue | Male | 88 |

Gradient Boost: ejemplo

Elegimos
 $Learning_Rate = 0.1$



| Height (m) | Favorite color | Gender | Weight (kg) |
|------------|----------------|--------|-------------|
| 1.6 | Blue | Male | 88 |



$$71.2 + (0.1 * 16.8) = 72.9$$

Gradient Boost: ejemplo

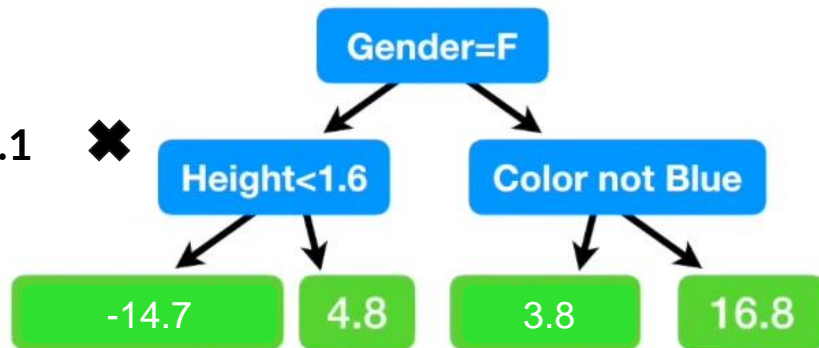
Elegimos
Learning_Rate = 0.1

72.9 no es tan buena predicción como 88, pero es mejor que la anterior: 71.2.

Se trata de un pequeño paso en la dirección correcta.

71.2

+ 0.1 ×



Segun **Jerome Friedman**, el autor de este método, la evidencia empírica sugiere que tomar muchos pasos pequeños en la dirección correcta resulta en una mejor predicción en un conjunto de datos de prueba.

Es decir: tendrá una **varianza baja**.

| Height (m) | Favorite color | Gender | Weight (kg) | Predicted |
|------------|----------------|--------|-------------|-----------|
| 1.6 | Blue | Male | 88 | 72.9 |



Gradient Boost: ejemplo

| Height (m) | Favorite color | Gender | Weight (kg) | Residual |
|------------|----------------|--------|-------------|----------|
| 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 | |

Generamos un nuevo árbol

Volvemos al Paso 2

Calculamos el error cometido por el estimador hasta ahora.

Residuo = (Peso observado - peso predicho)



Gradient Boost: ejemplo

| Residual |
|----------|
| 16.8 |
| 4.8 |
| -15.2 |
| 1.8 |
| 5.8 |
| -14.2 |

Residuos originales



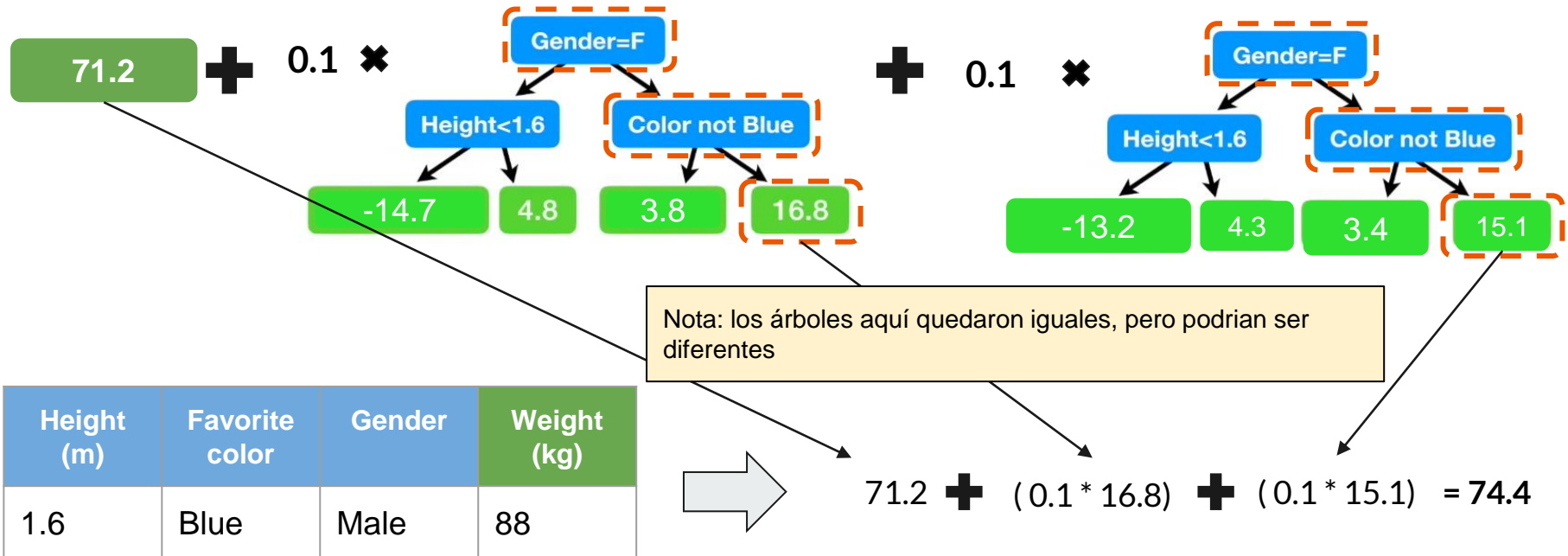
| Residual |
|----------|
| 15.1 |
| 4.3 |
| -13.7 |
| 1.4 |
| 5.4 |
| -12.7 |

Residuos nuevos



Podemos ver acá que todos los errores (residuos) se redujeron.
Dimos un paso en la dirección correcta

Gradient Boost: ejemplo





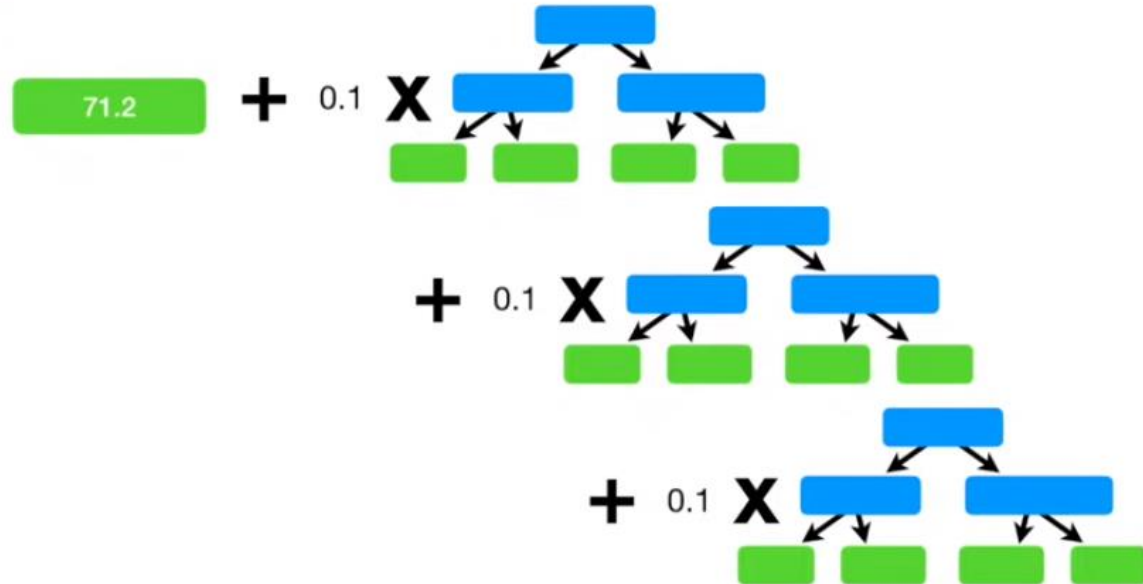
Gradient Boost: ejemplo

Seguimos añadiendo árboles, hasta que los residuos no cambien significativamente o bien alcanzamos un número preestablecido de árboles seteado como parámetro.

Finalmente para estimar un nuevo valor, cuando el método termina de entrenar, procedemos como antes, usando el valor inicial y sumando los residuos ponderados de cada árbol.

Gradient Boost: ejemplo

Con 3 árboles quedaría así



Video cómo este y muchos otros útiles



<https://www.youtube.com/channel/UCtYLUTgS3k1Fg4y5tAhLbw>

El autor es el Dr. Josh Starmer, profesor de la universidad de North Carolina