Ensambles de modelos

Ing Juan M. Rodríguez

La sabiduría de las multitudes

"The Wisdom of Crowds" James Surowiecki

En 1906 Galton (primo de Darwin) quería demostrar lo "bruta" que era la gente en base a un experimento.

Este constaba de adivinar el peso de una vaca en una exposición.

Cada participante debía poner su predicción en un buzón, aquel que más se acercarse se la llevaría a la casa.

La sabiduría de las multitudes

"The Wisdom of Crowds" James Surowiecki

Su hipótesis era que sólo podría ser bueno aquel que estuviese formado y tuviese experiencia en el rubro, el resto serían muy malos, tanto que aún promediando todas sus predicciones estarían muy lejos del valor real y del participante "sabio"

Sin embargo...

Ensambles de Modelos

- Entrenar varios modelos, c/u sobre datos distintos.
- Cada modelo *sobre-ajusta* de manera diferente.
  - Cada modelo: bajo sesgo, alta varianza.
  - Por ejemplo: árboles profundos.

Bias vs. Variance

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

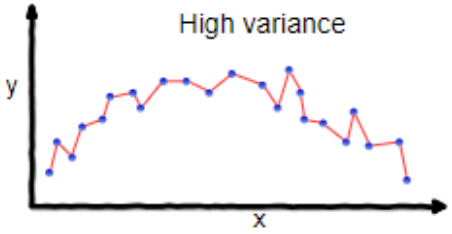
Variance (varianza)

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

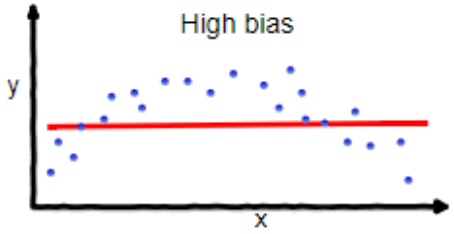
Un modelo con **varianza baja indica que cambiar los datos de entrenamiento produce cambios pequeños en la estimación**.

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***).

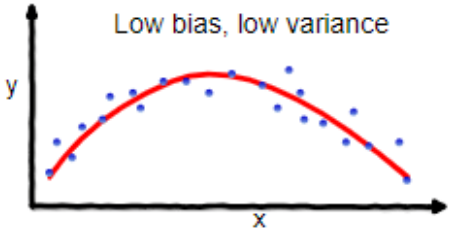
tradeoff bias vs variance



overfitting



underfitting

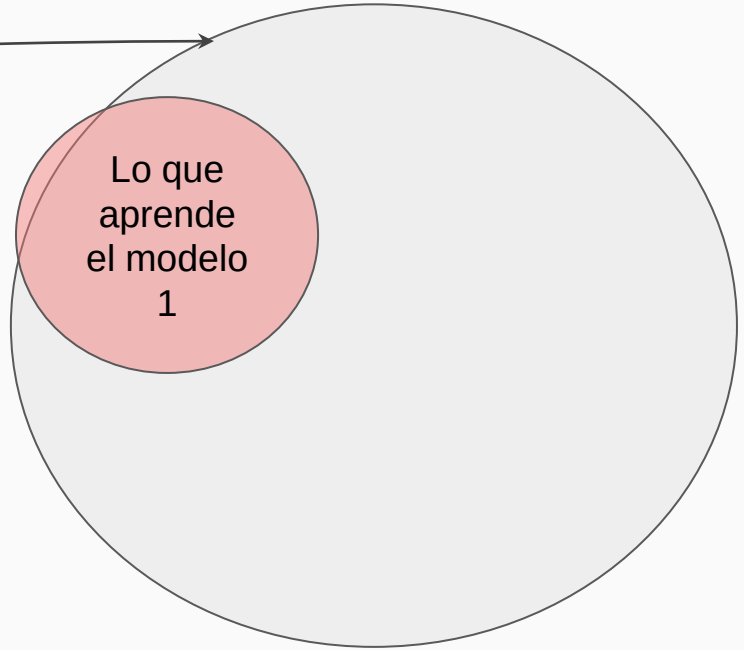


Good balance

Vuelta a ensambles

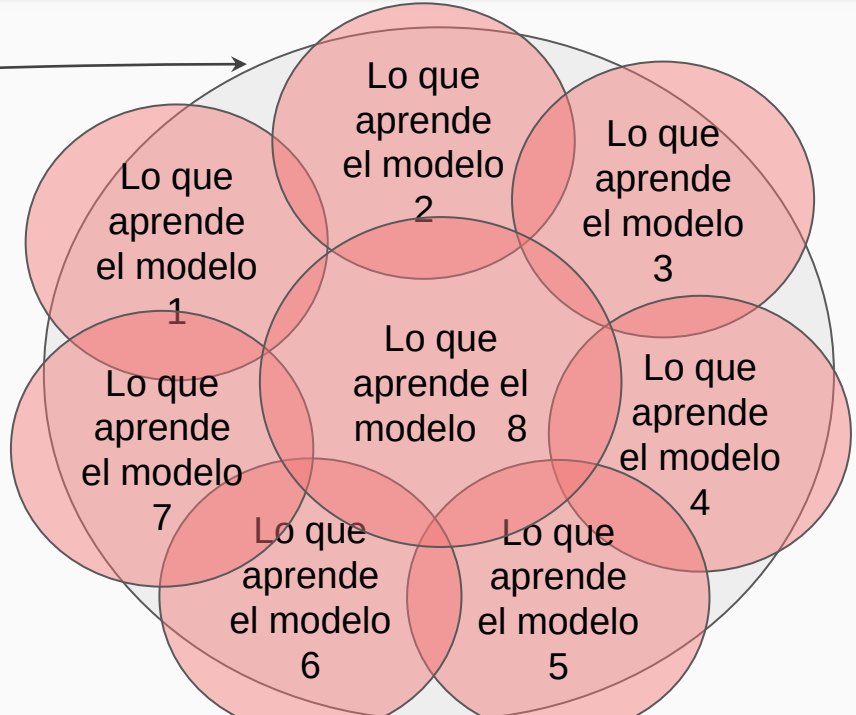
Ensambles de Modelos

El concepto a aprender

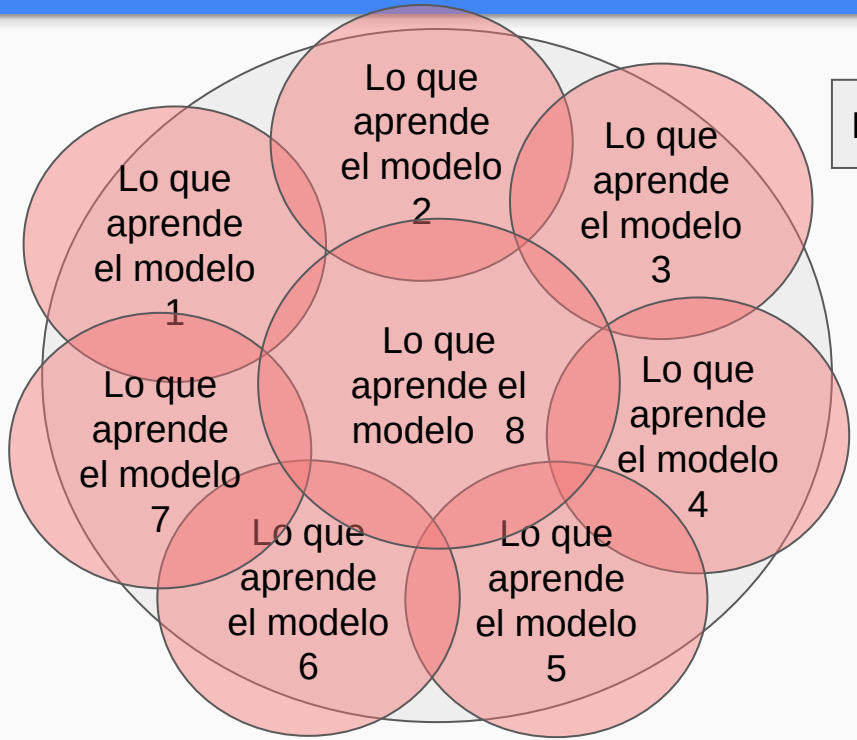


Ensambles de Modelos

El concepto a aprender



Ensambles de Modelos



Ensamble

Concepto

Ensambles de Modelos

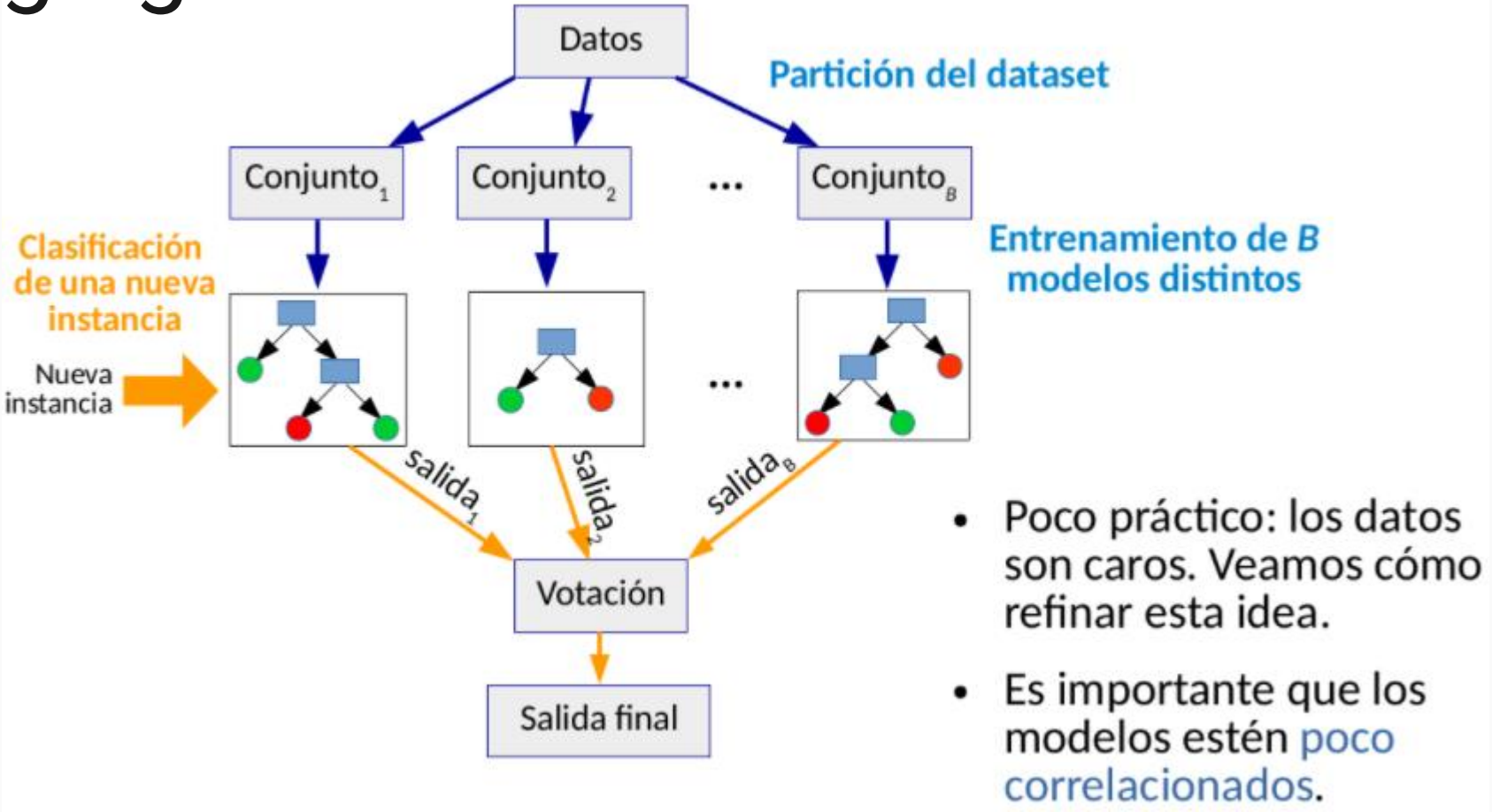
Votación: Para una nueva instancia, clasificarla con todos los modelos, y devolver la clase más elegida.

La votación reduce la varianza de la clasificación. (*Random Forest*)

Si los modelos individuales devuelven probabilidades, se puede hacer una votación ponderada.

Bagging

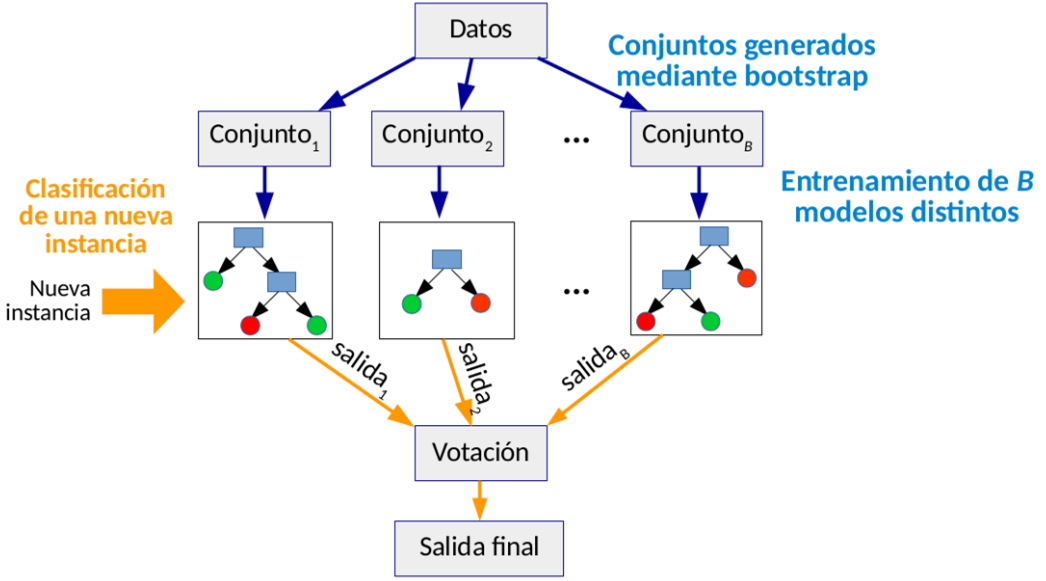
Bagging



- Poco práctico: los datos son caros. Veamos cómo refinar esta idea.
- Es importante que los modelos estén poco correlacionados.

Bagging

Es una técnica que consiste en construir nuevos conjuntos de entrenamiento usando ***bootstrap*** (muestras aleatorias con reemplazo) para entrenar distintos modelos, y luego combinarlos.



Bagging: paso a paso

1. Dividimos el conjunto de entrenamiento en distintos subconjuntos, obteniendo como resultado diferentes muestras aleatorias.
    a. Las muestras son uniformes (misma cantidad de individuos)
    b. Son muestras con reemplazo (los individuos pueden repetirse en el mismo conjunto de datos)
2. Entrenamos un modelo con cada subconjunto
3. Construimos un único modelo predictivo a partir de los anteriores

Bagging: características

- Disminuye la varianza en nuestro modelo final
- Muy efectivo en conjuntos de datos con varianza alta
- Puede reducir el ***overfitting***
- Puede reducir el ruido de los ***outliers*** (porque no aparecen en todos los datasets)
- Puede mejorar levemente con el voto ponderado

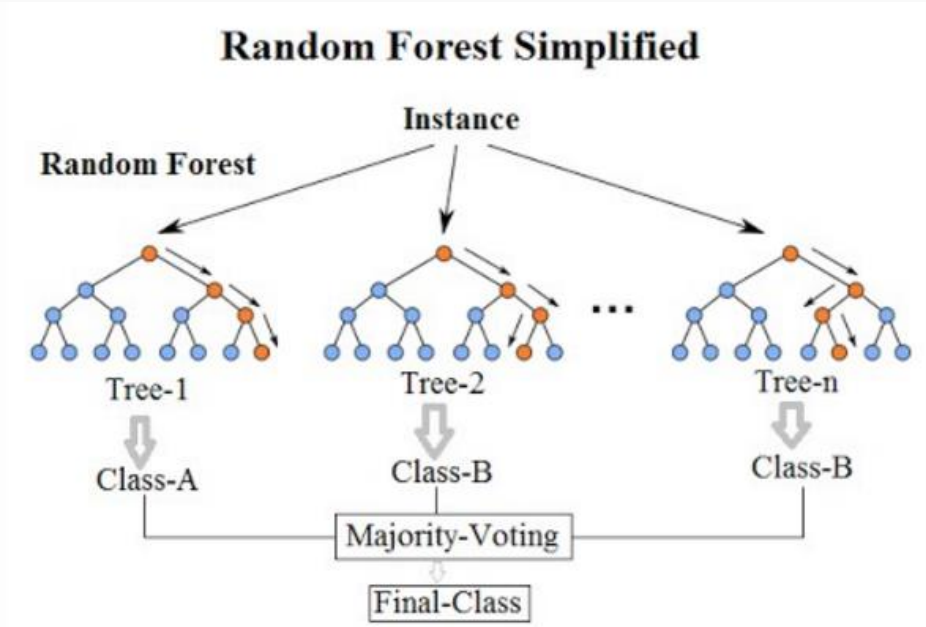
Bagging: problemas al usarlo con árboles

- Si pocos atributos son predictores fuertes, todos los árboles se van a parecer entre sí!

- Esos atributos terminarán cerca de la raíz, para todos los conjuntos generados con **bootstrap**.

Bagging: *Random Forest*

Igual a ***bagging*** tradicional, pero en cada nodo, considerar sólo un subconjunto de **m** atributos elegidos al azar.

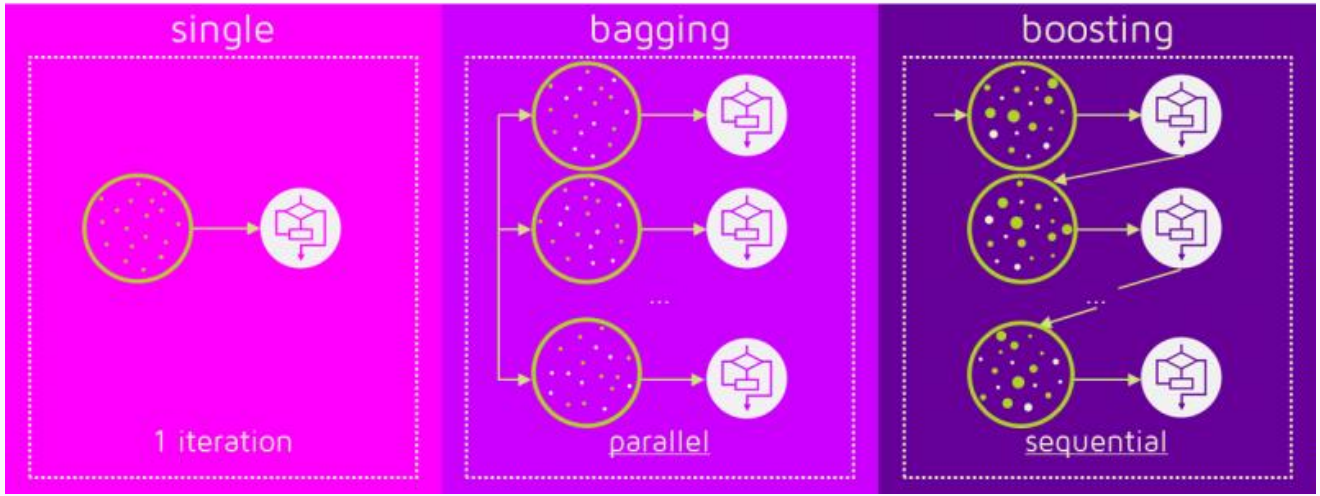


Boosting

Boosting

Alternativa a Bagging.

Buscar nuevos modelos para las instancias mal clasificadas por los anteriores.



Fuente: What is the difference between Bagging and Boosting? | Quantdare

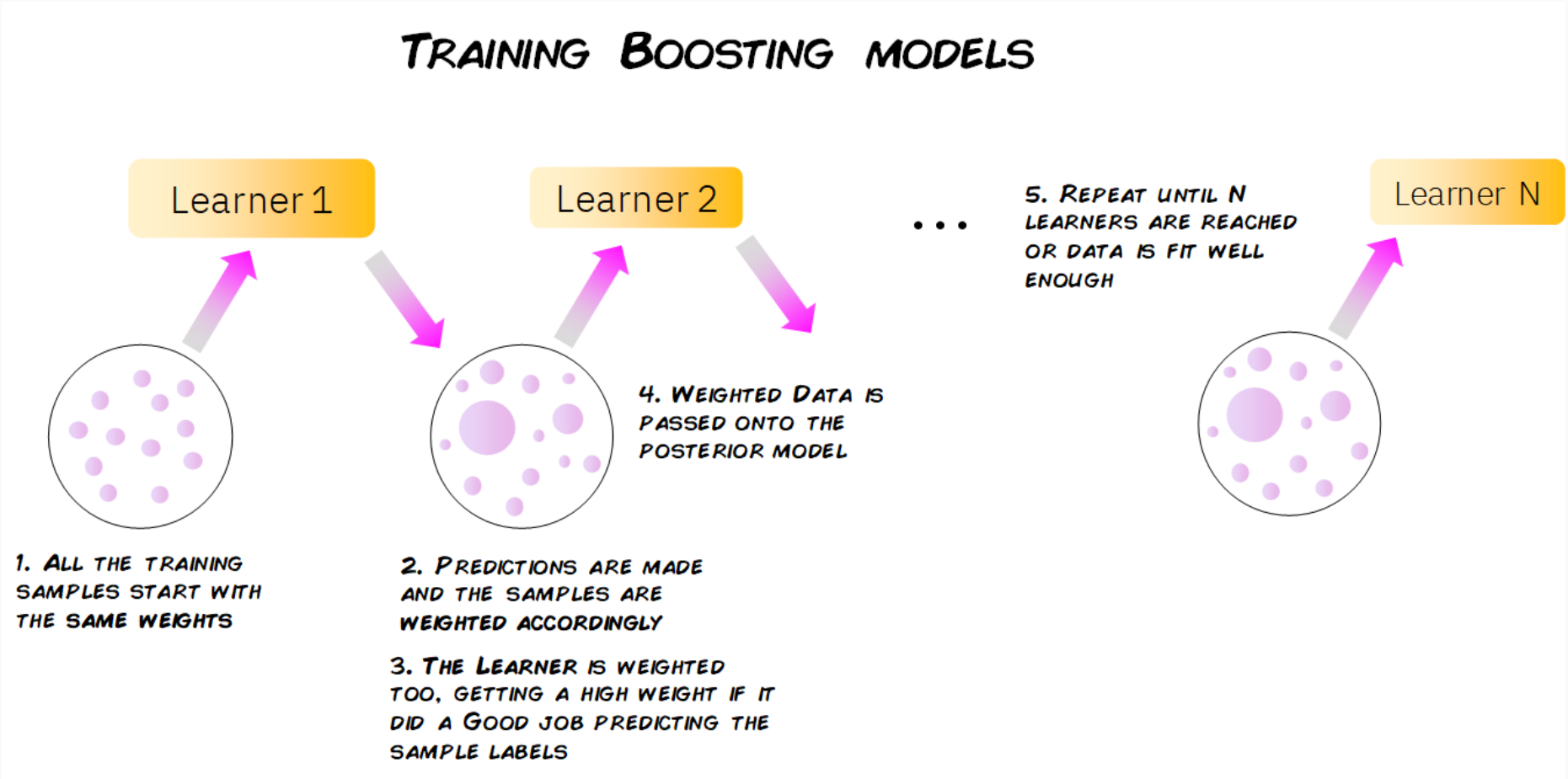
Boosting

1. Comenzar con un modelo (simple) entrenado sobre todos los datos: h_0

2. En cada iteración i, entrenar h_i dando mayor importancia a los datos mal clasificados por las iteraciones anteriores.

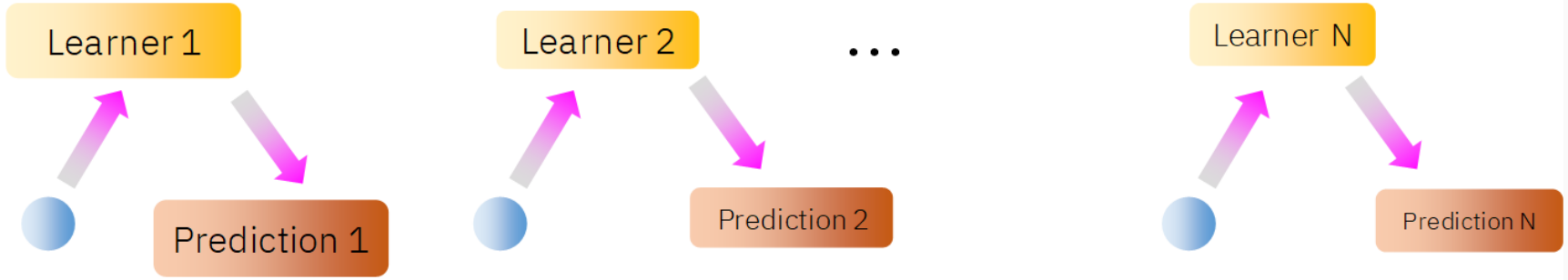
3. Terminar al conseguir cierto cubrimiento, o luego de un número de iteraciones.

4. Clasificar nuevas instancias usando una votación ponderada de todos los modelos construidos.



Fuente: https://towardsdatascience.com/what-is-boosting-in-machine-learning-2244aa196682

Predicting with Boosting models



1. The new sample is fed to every single model.

2. Each model makes its individual prediction, weighted depending on how well the model did during training.

3. The weighted predictions are combined to give a final output

Boosting: Modelos exitosos

- AdaBoost
- Gradient Boosting
- XGBoost: eXtreme Gradient Boosting

Boosting: XGBoost vs. Gradient Boosting

- La velocidad de entrenamiento de **XGBoost** es mucho menor gracias a su implementación y a estar mejor orientado al uso eficiente del hardware (GPU)

- El ***accuracy*** (o la métrica adecuada) también es mejor debido a que **XGBoost** maneja mejor el ***overfitting*** mediante regularizaciones (esto lo veremos más adelante)

Boosting resumen

- Necesita pesos, esto implica que:
  - Debemos adaptar algoritmo de aprendizaje
  - Y tomar muestras con reemplazo según pesos

- Puede sobreajustar