

# ¡Atrapa al Mejor Modelo!

## 1 ¿Qué ventajas ofrece combinar varios clasificadores mediante un sistema de votación?

Al considerar múltiples algoritmos básicos y combinarlos, reducimos la varianza de la predicción final, mejorando la capacidad de generalización de nuestro algoritmo final. Más aún, podemos alcanzar al usar algoritmos básicos y paralelización de sus entrenamientos, un modelo con un poder predictivo altamente competitivo.

## 2 Explica la diferencia conceptual entre votación dura y votación suave.

En votación dura, consideramos solo las predicciones concretas de los algoritmos del ensamble para obtener la predicción final, mientras que en la predicción suave consideramos las probabilidades, lo que nos permite tomar una decisión más fina sobre la predicción final, si asumimos que las probabilidades modeladas son fieles a las reales (ej. un algoritmo de 7 modelos con probabilidades 1.0, 1.0, 1.0, 0.49, 0.49, 0.49, 0.49 para una instancia, con votación dura clasificaría como 0, aún cuando los modelos en sí muestran a 3 completamente convencidos de que el resultado debe ser 1 y los otros 4 muestran una casi completa incertidumbre sobre la respuesta real).

## 3 ¿Qué tipo de problemas puede presentar una votación por mayoría simple?

Un problema podría ser el expuesto arriba, donde la votación por mayoría simple ignora la certeza concreta que tienen los estimadores sobre sus respuestas. Otro problema es que puede que los algoritmos terminen favoreciendo a una respuesta concreta en conjunto cuando, en general, habían otras respuestas que favorecían más, en las predicciones con segundas mayores probabilidades.

## 4. ¿En qué casos un ensamble puede tener peor desempeño que un modelo individual?

Cuando los modelos base son demasiado similares entre ellos, efectivamente dando lugar al mismo modelo entrenado muchas veces pero sobre un conjunto de datos más pequeño y por tanto, con un error real potencialmente mayor. Otro problema podría ser que los subconjuntos de muestra para entrenamiento de los modelos base sean muy pequeños para que los algoritmos puedan aprender realmente sin el riesgo de sobreajuste.

## 5. ¿Por qué es importante la diversidad entre los modelos base en un ensamble?

Porque, si todos los modelos fueran muy similares, el modelo final tras el ensamble sería esencialmente el mismo modelo al resultante tras haber entrenado directamente sobre los datos completos alguno de los modelos base. Cuando tenemos modelos diferentes, damos lugar a que la correlación entre ellos se disminuya y las ganancias en la generalización sean mayores.

## 6. Define el principio de independencia de alternativas irrelevantes y su relación con los sistemas de votación.

Indica que, si un votante prefiere a  $A$  sobre  $B$ , esa preferencia se debería mantener sin importar si consideramos un tercer elemento  $C$  o no dentro de los candidatos. Aunque suena razonable este principio, este puede fallar en casos como por ejemplo, cuando tenemos 2 candidatos  $A$  y  $B$  donde el clasificador selecciona a  $A$  con 0.60 y a  $B$  con 0.40, pero una vez que incluimos a  $C$ , todos los atributos que nos hacían querer seleccionar por  $A$  nos hacen querer seleccionar a  $C$  más, terminando con probabilidades de 0.25, 0.40, y 0.35 para  $A$ ,  $B$  y  $C$  respectivamente. Otro caso es que con votación por conteo de borda, puede no respetarse este principio.

## 7. ¿Qué diferencia existe entre una votación ponderada y una votación suave?

En una votación suave, todos los votantes se consideran iguales en su capacidad predictiva (es decir, que tanta razón pueden tener), y de este modo sus probabilidades se ponderan igual al momento de calcular la probabilidad final. En votación ponderada, agregamos a los votos un peso asociado a la exactitud de los modelos, de modo que los modelos con una mejor exactitud tengan mayor impacto en la decisión final, evitando que modelos con exactitudes bajas influyan mucho en el ensamble.

## 8. ¿Cómo afecta el sesgo y la varianza de los modelos base el resultado del ensamble?

Si todos los modelos base tienen sesgos similares, el sesgo del ensamble no va a cambiar mucho, aunque puede que empeore un poco al hacer los entrenamientos. La varianza sin embargo, se verá reducida, esto ocurre pues, si alguno de los modelos base fue entrenado con una muestra poco representativa, su impacto y su error se verá minimizado con el resto de modelos que sí fueron entrenados con muestras más representativas.

## 9. ¿Qué implicaciones tiene combinar modelos que tienden a cometer los mismos errores?

Que, al momento de considerar el ensamble entre estos, esos mismos errores se seguirán cometiendo, por ello es que buscamos una diversidad en los modelos base, de modo que los errores de unos modelos sean cancelados por los aciertos de los demás.

## 10. Explica por qué los métodos de ensamble no contradicen el teorema del no free lunch

El teorema del NFL nos indica que no existe un algoritmo de aprendizaje que sea universalmente bueno para todos los casos y problemas. Con los métodos de ensamble como bagging somos capaces de reducir la varianza, sí, pero que tanta varianza podemos reducir va a depender de que tan correlacionadas van a ser las submuestras que tomaremos para los modelos base, y que tan buenos son los modelos base en sí para alcanzar el error mínimo posible, lo cual también va a depender de la complejidad de los modelos (y su relación al tamaño de la muestra). Si no se tienen condiciones favorables en lo anterior, los modelos de ensamble seguirán teniendo problemas de generalización muy grandes, y serán superados por elecciones de modelos más adecuadas a cada problema específico.