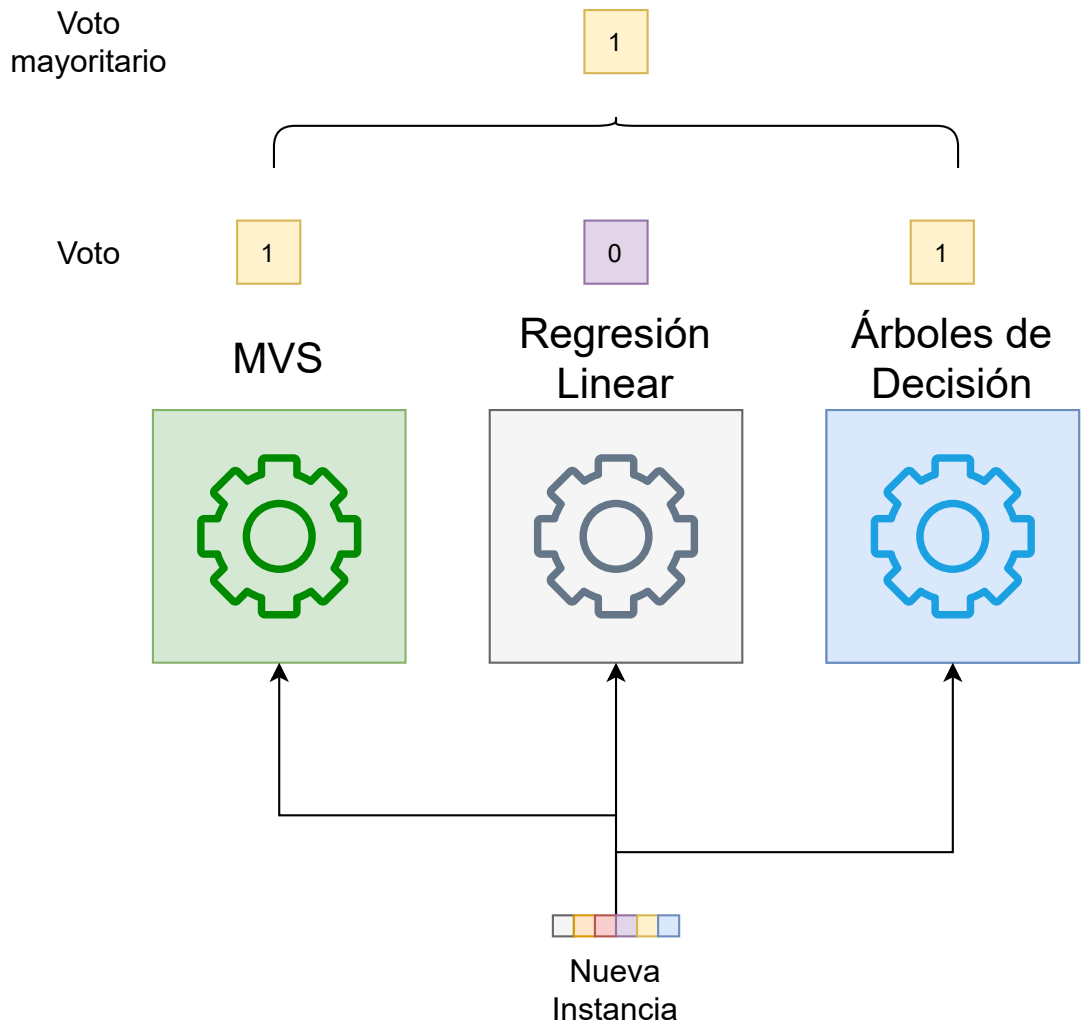
Ensemble Learning

Bagging y Pasting

Luis Norberto Zúñiga Morales

12 de septiembre de 2022

En clases anteriores...



En clases anteriores...

Pregunta

¿Se les ocurre alguna otra forma de jugar con los elementos de un ensamble para mejorar su rendimiento?

Bagging

- En el caso del Voting Classifier, se combinan diferentes opiniones por medio de distintos modelos de aprendizaje.

Bagging

- En el caso del Voting Classifier, se combinan diferentes opiniones por medio de distintos modelos de aprendizaje.
- Cada uno aporta sus puntos de vista, se agregan y se elige un resultado final por medio de un voto duro o suave, o un promedio para el caso de regresión.

Bagging

- En el caso del Voting Classifier, se combinan diferentes opiniones por medio de distintos modelos de aprendizaje.
- Cada uno aporta sus puntos de vista, se agregan y se elige un resultado final por medio de un voto duro o suave, o un promedio para el caso de regresión.
- ¿Qué pasa si jugamos con el conjunto de datos, en lugar de los modelos de entrenamiento?

Bagging

- En 1995, Leo Breiman [1] introdujo la idea del Bagging.

Bagging

- En 1995, Leo Breiman [1] introdujo la idea del Bagging.
- *Bagging* = *Bootstrap Aggregation*.

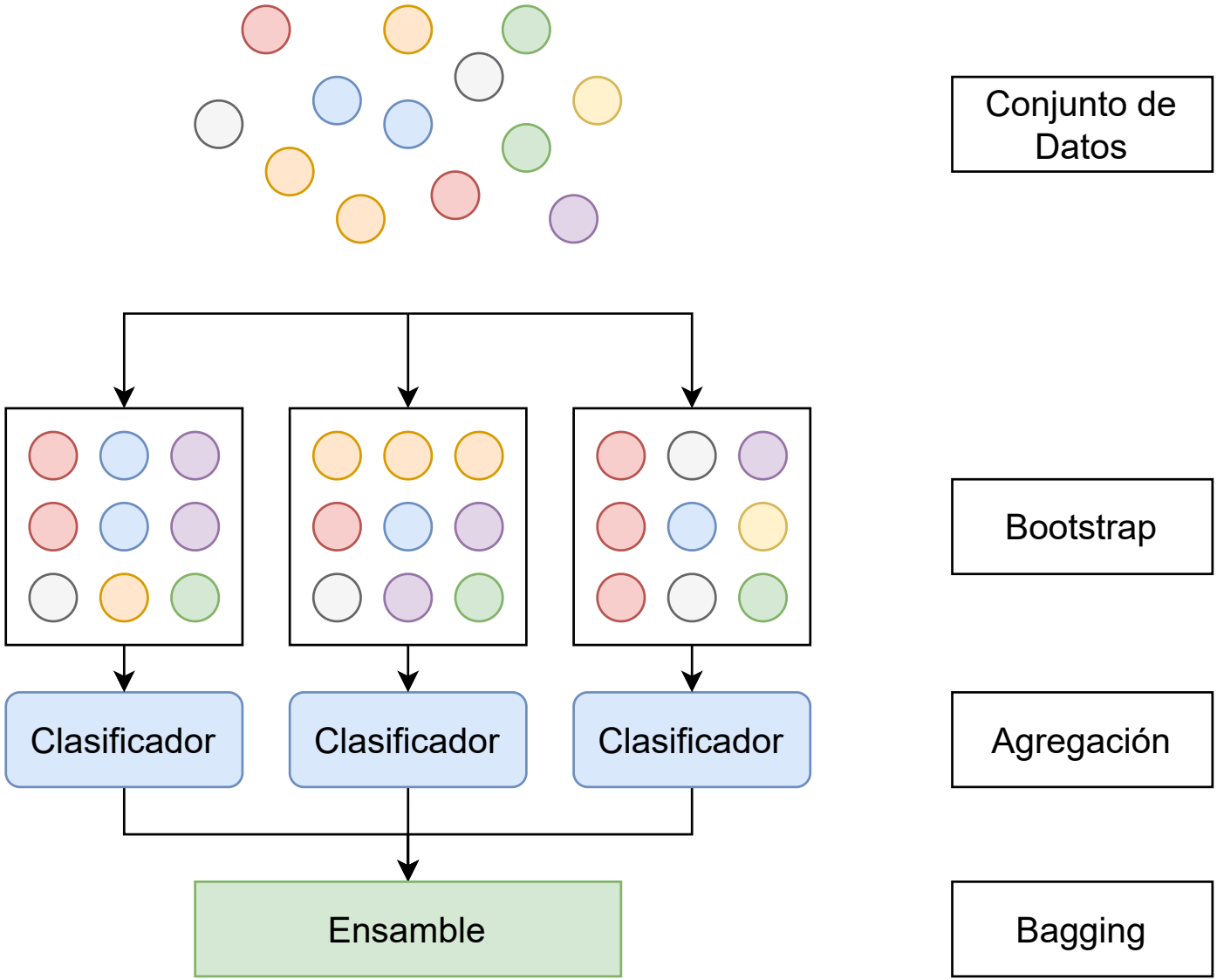
Bagging

- En 1995, Leo Breiman [1] introdujo la idea del Bagging.
- *Bagging* = *Bootstrap Aggregation*.
- Consiste en eligir varias muestras del conjunto de datos con remplazo.

Bagging

- En 1995, Leo Breiman [1] introdujo la idea del Bagging.
- *Bagging* = *Bootstrap Aggregation*.
- Consiste en eligir varias muestras del conjunto de datos con remplazo.
- Cada muestra se entrena de forma independiente y, dependiendo de la tarea (regresión o clasificación), se elige el promedio o mayoría de las predicciones.

Bagging


Conjunto de Datos
Bootstrap
Clasificador
Clasificador
Clasificador
Agregación
Ensamble
Bagging

Bagging

- Un conjunto de entrenamiento L consiste de puntos de datos $\{(\boldsymbol{x}_i, y_i), i = 1, ..., N\}$.

Bagging

- Un conjunto de entrenamiento L consiste de puntos de datos $\{(\boldsymbol{x}_i, y_i), i = 1, ..., N\}$.
- Supongamos que se emplea un modelo de aprendizaje para encontrar un función predictora $y = f(\boldsymbol{x}, L)$.

Bagging

- Un conjunto de entrenamiento L consiste de puntos de datos $\{(\boldsymbol{x}_i, y_i), i = 1, ..., N\}$.
- Supongamos que se emplea un modelo de aprendizaje para encontrar un función predictora $y = f(\boldsymbol{x}, L)$.
- Supongamos que se tiene una secuencia de conjuntos de entrenamiento $\{L_k\}$ cada uno con N observaciones independientes del conjunto original L.

Bagging

- Un conjunto de entrenamiento L consiste de puntos de datos $\{(\boldsymbol{x}_i, y_i), i = 1, ..., N\}$.
- Supongamos que se emplea un modelo de aprendizaje para encontrar un función predictora $y = f(\boldsymbol{x}, L)$.
- Supongamos que se tiene una secuencia de conjuntos de entrenamiento $\{L_k\}$ cada uno con N observaciones independientes del conjunto original L.
- El objetivo del Bagging es usar $\{L_k\}$ para obtener una mejor función predictora $\tilde{f}(\boldsymbol{x}, L_k)$ mediante la construcción de nuevas funciones $\{\phi_{L_k} | k = 1, ..., M\}$.

Bagging

Para el caso de un problema de clasificación, se puede considerar el voto duro:

$$\psi_{\phi_{L_1},\ldots,\phi_{L_M}}(\boldsymbol{x}) = \underset{c \in y}{\operatorname{argmax}} \sum_{m=1}^{M} 1(\phi_{L_m}(\boldsymbol{x}) = c) \tag{1}$$

o el voto suave:

$$\psi_{\phi_{L_1},\ldots,\phi_{L_M}}(\boldsymbol{x}) = \underset{c \in y}{\operatorname{argmax}} \frac{1}{M} \sum_{m=1}^{M} \hat{p}_{L_m}(Y = c | X = \boldsymbol{x}) \tag{2}$$

Bagging

Para el caso de regresión, se puede optar por un promedio de los valores numéricos que arroja el modelo de aprendizaje:

$$\psi(\boldsymbol{x}_i) = \frac{1}{M} \sum_{k=1}^{M} \phi_{L_k}(\boldsymbol{x}_i) \tag{3}$$

Bagging

Example (Bagging)

Sea $L = \{1, 2, 3, 4, 5, 6\}$. Al realizar el muestro con reemplazo, se pueden obtener los siguientes subconjuntos para entrenar nuevos modelos de aprendizaje:

$$L_1 = 1, 2, 2, 3, 4, 1$$
$$L_2 = 1, 5, 6, 2, 2, 3$$
$$L_3 = 2, 3, 1, 5, 6, 6$$

Después, se crean nuevos modelos independientes $\phi_{L_1}, \phi_{L_2}, \phi_{L_3}$ que se entrenan por separado y se someten a un voto duro o suave para el caso de clasificación, o un promedio para el caso de regresión.

Bagging

La idea de generar nuevos subconjuntos de datos a partir del conjunto de datos original tiene propiedades atractivas desde el punto de vista estadístico:

- Las muestras se obtienen de una fuente que, en teoría, representa la realidad y sigue la distribución de probabilidad que rige el fenómeno observado.

Bagging

La idea de generar nuevos subconjuntos de datos a partir del conjunto de datos original tiene propiedades atractivas desde el punto de vista estadístico:

- Las muestras se obtienen de una fuente que, en teoría, representa la realidad y sigue la distribución de probabilidad que rige el fenómeno observado.
- Las muestras son independientes entre ellas, lo cual lleva a suponer que son muestras independientes e idénticamente distribuidas.

Bagging

No todo es perfecto. El conjunto de datos original debe ser lo suficientemente grande para que:

- Pueda capturar la complejidad de la distribución de probabilidad y el nuevo muestreo pueda equipararse a uno de la distribución original.

Bagging

No todo es perfecto. El conjunto de datos original debe ser lo suficientemente grande para que:

- Pueda capturar la complejidad de la distribución de probabilidad y el nuevo muestreo pueda equipararse a uno de la distribución original.
- Para que las muestras no presenten correlación entre ellas y el conjunto de datos.

Bagging

En resumen...

Debe existir representatividad e independencia.

Bagging

El algoritmo consta de los siguientes pasos:

1. **Bootstrapping**: se realizan múltiples muestras o subconjuntos del conjunto de entrenamiento. Dichas muestras se realizan mediante muestreo con reemplazo.

Bagging

El algoritmo consta de los siguientes pasos:

1. **Bootstrapping**: se realizan múltiples muestras o subconjuntos del conjunto de entrenamiento. Dichas muestras se realizan mediante muestreo con reemplazo.
2. **Entrenamiento**: cada muestra o subconjunto se entrena de forma paralela e independiente entre ellas, ya sea con modelos débiles o modelos base.

Bagging

El algoritmo consta de los siguientes pasos:

1. **Bootstrapping**: se realizan múltiples muestras o subconjuntos del conjunto de entrenamiento. Dichas muestras se realizan mediante muestreo con reemplazo.
2. **Entrenamiento**: cada muestra o subconjunto se entrena de forma paralela e independiente entre ellas, ya sea con modelos débiles o modelos base.
3. **Agregación**: Según la tarea, se toma un promedio o la mayoría de las predicciones para calcular una estimación más precisa.

¿Dónde se puede utilizar Bagging?

Breiman menciona que usar Bagging es efectivo en modelos de aprendizaje inestables, donde pequeños cambios en el conjunto de datos causan grandes cambios en las predicciones del modelo.

Ejemplos de dichos modelos inestables incluyen:

- Árboles de Decisión
- Redes Neuronales Artificiales
- Algunos modelos de Regresión

Bagging

Dentro de las ventajas que supone utilizar Bagging se encuentran las siguientes:

- Reduce el factor de la varianza cuando un modelo de aprendizaje débil presenta bajo sesgo y alta varianza.

Bagging

Dentro de las ventajas que supone utilizar Bagging se encuentran las siguientes:

- Reduce el factor de la varianza cuando un modelo de aprendizaje débil presenta bajo sesgo y alta varianza.
- El ensamble mejora el rendimiento de los modelos base débiles.

Bagging

Dentro de las ventajas que supone utilizar Bagging se encuentran las siguientes:

- Reduce el factor de la varianza cuando un modelo de aprendizaje débil presenta bajo sesgo y alta varianza.
- El ensamble mejora el rendimiento de los modelos base débiles.
- Se puede realizar el entrenamiento en paralelo.

Bagging

Sin embargo, también presenta algunas desventajas:

- Para modelos de aprendizaje débiles con alto sesgo, al emplear Bagging se transfiere el alto sesgo. En general, el Bagging no reduce el sesgo, únicamente la varianza.

Bagging

Sin embargo, también presenta algunas desventajas:

- Para modelos de aprendizaje débiles con alto sesgo, al emplear Bagging se transfiere el alto sesgo. En general, el Bagging no reduce el sesgo, únicamente la varianza.
- Pérdida de la interpretabilidad de los modelos de aprendizaje empleados.

Bagging

Sin embargo, también presenta algunas desventajas:

- Para modelos de aprendizaje débiles con alto sesgo, al emplear Bagging se transfiere el alto sesgo. En general, el Bagging no reduce el sesgo, únicamente la varianza.
- Pérdida de la interpretabilidad de los modelos de aprendizaje empleados.
- Puede ser complejo computacionalmente, ya que los subconjuntos pueden llegar a ser de gran tamaño, dependiendo del conjunto de datos base que se utilice.

Bagging

Tarea

1. Leer el artículo de Breiman [1] sobre el Bagging.
2. Prestar atención al algoritmo propuesto que utiliza Árboles de Decisión.
3. Para el próximo miércoles: Realizar su propio programa para realizar Bagging.
   - Realizar el muestreo aleatorio con reemplazo.
   - Entrenar cada muestra generada con modelos base.
   - La agregación se realiza con voto duro.

Pasting

- En 1999, Breiman [2] introdujo el algoritmo de Pasting como una alternativa de EL.

Pasting

- En 1999, Breiman [2] introdujo el algoritmo de Pasting como una alternativa de EL.
- El conjunto de entrenamiento L es muy grande para guardarlo en la memoria de la computadora.

Pasting

- En 1999, Breiman [2] introdujo el algoritmo de Pasting como una alternativa de EL.
- El conjunto de entrenamiento L es muy grande para guardarlo en la memoria de la computadora.
- La idea del Pasting es crear muestras de menor tamaño de L que sirven de entrada a distintos modelos de aprendizaje cercanos al óptimo.

Pasting

Ideas clave para la formulación del algoritmo:

- Supongamos que hasta el momento se han construido k predictores.

Pasting

Ideas clave para la formulación del algoritmo:

- Supongamos que hasta el momento se han construido k predictores.
- Un nuevo conjunto de entrenamiento de tamaño N se selecciona de L, ya sea por muestreo aleatorio (usualmente sin reemplazo) o de importancia.

Pasting

Ideas clave para la formulación del algoritmo:

- Supongamos que hasta el momento se han construido k predictores.
- Un nuevo conjunto de entrenamiento de tamaño N se selecciona de L, ya sea por muestreo aleatorio (usualmente sin reemplazo) o de importancia.
- El $(k+1)$-ésimo predictor se entrena en este nuevo conjunto de entrenamiento y se agrega con los k anteriores.

Pasting

Ideas clave para la formulación del algoritmo:

- Supongamos que hasta el momento se han construido k predictores.
- Un nuevo conjunto de entrenamiento de tamaño N se selecciona de L, ya sea por muestreo aleatorio (usualmente sin reemplazo) o de importancia.
- El $(k+1)$-ésimo predictor se entrena en este nuevo conjunto de entrenamiento y se agrega con los k anteriores.
- La agregación se realiza por voto mayoritario (sin pesos).

Pasting

Pasting

- Si se utiliza un muestreo aleatorio para seleccionar el conjunto de entrenamiento, este se llama *Rprecinct* y el procedimiento se llama *pasting Rvotes* (R = aleatorio).
- Si se realiza mediante muestreo por importancia, el conjunto de entrenamiento se llama *Iprecint* y el procedimiento, *pasting Ivotes* (I = importancia).

Pasting

Ideas clava para su evaluación y condición de paro:

- Se actualiza una estimación $e(k)$ del error de generalización para la k-ésima agregación $e(k)$.

Pasting

Ideas clava para su evaluación y condición de paro:

- Se actualiza una estimación $e(k)$ del error de generalización para la k-ésima agregación $e(k)$.
- El pasting se detiene cuando $e(k)$ deja de disminuir.

Pasting

Ideas clava para su evaluación y condición de paro:

- Se actualiza una estimación $e(k)$ del error de generalización para la k-ésima agregación $e(k)$.
- El pasting se detiene cuando $e(k)$ deja de disminuir.
- La estimación de $e(k)$ se puede obtener usando un conjunto de prueba fijo, aunque también puede usarse el conjunto fuera de la bolsa (*out-of-bag*).

Out-of-the-Bag

- Tanto en el Bagging como en el Pasting, algunas instancias pueden repetirse múltiples veces durante el muestreo al momento de crear los subconjuntos que se utilizan para entrenar los nuevos modelos del ensamble.

Out-of-the-Bag

- Tanto en el Bagging como en el Pasting, algunas instancias pueden repetirse múltiples veces durante el muestreo al momento de crear los subconjuntos que se utilizan para entrenar los nuevos modelos del ensamble.
- Es posible que otros puntos no lleguen a ser seleccionados ni una sola vez durante el proceso.

Out-of-the-Bag

- Tanto en el Bagging como en el Pasting, algunas instancias pueden repetirse múltiples veces durante el muestreo al momento de crear los subconjuntos que se utilizan para entrenar los nuevos modelos del ensamble.
- Es posible que otros puntos no lleguen a ser seleccionados ni una sola vez durante el proceso.
- Durante el Bagging, que considera muestreo con reemplazo, alrededor del 63 % de las instancias del conjunto de entrenamiento son seleccionadas durante el muestreo.

Out-of-the-Bag

- Tanto en el Bagging como en el Pasting, algunas instancias pueden repetirse múltiples veces durante el muestreo al momento de crear los subconjuntos que se utilizan para entrenar los nuevos modelos del ensamble.
- Es posible que otros puntos no lleguen a ser seleccionados ni una sola vez durante el proceso.
- Durante el Bagging, que considera muestreo con reemplazo, alrededor del 63 % de las instancias del conjunto de entrenamiento son seleccionadas durante el muestreo.
- El restante 37 % de las instancias que no se eligen conforman el conjunto fuera de la bolsa (*out-of-bag*).

Pasting

Tarea

1. Leer el artículo de Breiman [2] sobre el Pasting.
2. Para el próximo miércoles: Elegir un conjunto de datos grande y utilizar Pasting para entrenarlo.

Bibliografía

[1] Leo Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[2] Leo Breiman. Pasting small votes for classificationin large databases and on-line. *Machine Learning*, 36:85–103, 1999.