

Sistemas Inteligentes

Sesion 4/11

November 4, 2021

1 Introducción

Tanto Boosting como Bagging se tratan de algoritmos de metálgortmos de aprendizaje automático. Estos algoritmos funcionan mediante la combinación de clasificadores débiles para hacer un juicio de valor sobre un conjunto de datos de entrada. Estos clasificadores débiles se unen para formar unos clasificadores más potentes, también se les conoce como clasificadores fuertes, se puede decir que la unión hace la fuerza.

Cada uno de estos dos algoritmos tiene unas ventajas y desventajas que se verán a continuación.

2 Bagging

Los métodos de bagging son métodos donde los algoritmos más simples son usados en paralelo, el principal objetivo de esto es aprovecharse de la independencia que existe entre los algoritmos más simples, ya que podemos reducir el error al promediar las salidas de los modelos más simples. En otras palabras, elegiremos lo que elija la mayoría. Bagging usará votación para métodos de *clasificación* y *promedio* para métodos de regresión.



Figure 1: Métodos paralelos de Bagging

3 Boosting

Por su parte Boosting (por ejemplo AdaBoost, que se implementará en prácticas), estos algoritmos tienen como objetivo principal la reducción del sesgo. En los métodos de Boosting, los modelos

simples (clasificadores débiles) se emplean de manera secuencial, esto quiere decir, que el modelo $i+1$ depende del modelo anterior. La ventaja de esto es que nos aprovechamos de la dependencia que existe entre los modelos más simples. Los clasificadores débiles introducen el concepto de confianza, en cada iteración esta confianza se actualiza, aumentando en los clasificadores que mejor clasifiquen los ejemplos.

La distribución de probabilidad es la misma para todos al principio, pero en futuras iteraciones va variando, como se ha comentado.

Se puede mejorar el rendimiento de estos algoritmos si en los clasificadores débiles posteriores le damos importancia a los ejemplos más complicados de clasificar.

El juicio de valor de cada clasificador débil será mediante *votación* si estamos frente a problemas de clasificación y mediante *suma ponderada* si es un problema de regresión.



Figure 2: Métodos secuenciales de Boosting

4 Diferencias

Como se ha visto en los apartados anteriores, la gran diferencia es como tratan los clasificadores ambos métodos, por su parte en Bagging los clasificadores débiles al enunciar el juicio de valor todos tienen el mismo peso, mientras que en Boosting los que más peso tendrán son los que más confianza tengan.