# Descripción de funcionamiento del algoritmo BoostAODE

El algoritmo se basa en el funcionamiento del algoritmo AdaBoost, y utilizando una serie de hiperparámetros se activan cada una de las posibles alternativas.

Los hiperparámetros que están definidos en el algoritmo son:

* repeatSparent (boolean): Permite que se repitan variables del dataset como padre de un SPODE. Valor por defecto falso.
* maxModels (int): número máximo de modelos (SPODEs) a construir. Este hiperparámetro únicamente es tenido en cuenta si “repeatSparent” se establece como verdadero. Valor por defecto 0.
* ascending (boolean): Establece el orden (ascendente/descendente) en el que se procesarán las variables del dataset para elegir los padres de los SPODEs. Valor por defecto false.
* convergence (boolean): Establece si se utilizará la convergencia del resultado como condición de finalización. En caso de establecer este hiperparámetro como verdadero, el conjunto de datos de entrenamiento que se pasa al modelo se divide en dos conjuntos uno que servirá como datos de entrenamiento y otro de test (por lo que la partición de test original pasará a ser en este caso una partición de validación). La partición se realiza tomando como la primera partición generada por un proceso de generación de 5 particiones estratificadas y con una semilla prefijada. La condición de salida es que la diferencia entre la precisión obtenida por el modelo actual, y la obtenida por el modelo anterior sea mayor que 1e-4, en caso contrario se sumará uno al número de modelos que empeora el resultado (ver siguiente hiperparámetro). Valor por defecto false.
* tolerance (int): Establece el número máximo de modelos que pueden empeorar el resultado sin suponer una condición de finalización. Valor por defecto 0.
* select\_features ({“IWSS”, “FCBF”, “CFS”, “”}): Selecciona en caso de establecerlo, el método de selección de variables que se utilizará para construir unos modelos iniciales para el ensemble que se incluirán sin tener en cuenta ninguna de las otras condiciones de salida. Estos modelos tampoco actualizan o utilizan los pesos que utiliza el algoritmo de Boosting y su significancia se establece en 1.
* threshold (double): Establece el valor necesario para los algoritmos IWSS y FCBF para poder funcionar. Los valores aceptados son:
  + IWSS: [0, 0.5]
  + FCBF: [1e-7, 1]

El algoritmo realiza los siguientes pasos:

* Si se ha establecido select\_features se crean tantos SPODEs como variables selecciona el algoritmo correspondiente y se marcan como utilizadas estas variables
* Se establecen los pesos iniciales de los ejemplos como 1/m
* Bucle principal de entrenamiento
  + Ordena las variables por orden de información mutua con la variable clase y se procesan en orden ascente o descendente según valor del hiperparámetro
  + Si no se ha establecido la repetición de padres, se marca la variable como utilizada
  + Crea un Spode utilizando como padre la variable seleccionada
  + Entrena el modelo y calcula la variable clase correspondiente al dataset de entenamiento
  + Actualiza los pesos asociados a los ejemplos de acuerdo a la exresión:

(si el ejemplo ha sido mal clasificado)

(si el ejemplo ha sido bien clasificado)

* + Establece la significancia del modelo como αt
  + Si se ha establecido el hiperparámetro de convergencia, se calcula el valor de la precisión sobre el dataset de test que hemos separado en un paso inicial
  + Condiciones de salida:
    - εt > 0,5 => se penaliza a los ejemplos mal clasificados
    - Número de modelos con precisión peor mayor que tolerancia y convergencia establecida
    - No hay más variables para crear modelos y no repeatSparent
    - Número de modelos > maxModels si repeatSparent está establecido