

# Tema 6 Boosting y Adaboost

El aprendizaje supervisado es una técnica para obtener una función a partir de datos de entrenamiento. Los datos de entrenamiento consisten en pares en los cuales un elemento son los datos de entrada y el otro los resultados deseados. Las funciones que se obtengan serán aquellas que compondrán el sistema que estamos entrenando. La obtención de clasificadores puede hacerse mediante la combinación de clasificadores débiles, estos deben de ser simples y deben funcionar por lo menos mejor que una clasificación aleatoria, es decir, deben ser moderadamente precisas. Estos pueden combinarse mediante Bagging y boosting.

Este método ayuda a mejorar los clasificadores inestables, como redes neuronales o árboles de decisión (pequeños cambios en el conjunto de entrenamiento llevan a diferentes clasificadores y grandes cambios en el porcentaje de aciertos).

En el boosting se hace uso del muestreo ponderado, este consiste en, en lugar de hacer un muestreo aleatorio de los datos de entrenamiento, se ponderan las muestras para concentrar el aprendizaje en los ejemplos más difíciles. Intuitivamente, los ejemplos más cercanos a la frontera de decisión son más difíciles de clasificar, y recibirán pesos más altos.

Por otro lado, el bagging, consiste en lugar de combinar los clasificadores con el mismo peso en el voto, se usa un voto ponderado. Esta es la regla de combinación para el conjunto de clasificadores débiles. En conjunción con la estrategia de muestreo anterior, esto produce un clasificador más fuerte.