

Lista de Exercícios #8

Aprendizagem em conjunto e florestas aleatórias

1. Sim. É possível agrupar os resultados da predição dos diversos classificadores e aplicar o classificador do tipo *Hard Voting*, que assume como resultado, a predição que obtiver o maior número de votos.
2. A votação rígida é a média das predições dos classificadores, enquanto a votação suave, leva em consideração a probabilidade de cada classe, ou seja, é dado maior peso para a classe com maior confiança.
3. Sim. É possível realizar o treinamento das classes em paralelo, onde a predição é executada em diversas CPUs ou servidores.
4. O emprego do *Out of Bag* possibilita que um *dataset* que não tenha sido utilizado no treinamento, ou seja, não é conhecido pelo componente que realiza a predição, seja utilizado na validação de uma floresta aleatória. Desta forma, não há necessidade de um set de validação separado ou uso de validação cruzada para avaliar o desempenho de uma árvore.
5. Ao ampliar uma árvore utilizando *Random Forest*, a cada nó, é considerado um *subset* aleatório de limiar para realizar a divisão. Ao contrário da *Random Forest*, as árvores extras usam um limiar aleatório para cada *feature*, o que a torna ainda mais aleatória. As árvores extras são muito mais rápidas de serem treinadas do que as árvores aleatórias normais, já que não há a necessidade de calcular o melhor limiar para cada *feature* em cada nó.