

Aluno(a): **Mayomona Lando Filipe** _____.

Matrícula: **836**.

Aprendizagem em conjunto e florestas aleatórias

1. Digamos que você treinou cinco modelos diferentes com exatamente os mesmos dados de treinamento e todos alcançam 95% de precisão, existe alguma chance de você poder combinar esses modelos para obter melhores resultados? Se sim, como? Se não, por que?

R: Você pode combinar esses modelos para obter melhores resultados sim. Será o seguinte: Crie-se um classificador por meio de um nível majoritário, um classificador de votação, porque o classificador tem maior precisão e melhor desempenho, ou seja, você pode tentar mesclá-los em um conjunto de votação, o que geralmente Traga melhores resultados. É ainda melhor se eles forem treinados em diferentes exemplos de treinamento (esse é o objetivo de ensacamento e colagem), mas se eles não forem treinados, ainda funcionará enquanto o modelo for muito diferente.

2. Qual é a diferença entre classificadores de votação rígida e suave?

R:A diferença é que, na votação Rígida, cada classificador vota em uma classe e o voto da maioria vence. Em outras palavras, esse classificador conta apenas o número de votos para cada classificador na coleção e seleciona a categoria com o maior número de votos. Na votação suave, cada classificador individual fornece um valor de probabilidade de que um ponto de dados específico pertença a uma categoria de destino específica. Em seguida, a tag de destino com a maior soma de chances ponderadas ganhará a votação. Ou seja, o classificador de votação suave calcula a probabilidade média estimada da categoria para cada categoria e seleciona a categoria com a maior probabilidade.

3. É possível acelerar o treinamento de um **bagging ensemble** distribuindo-o por vários servidores? E quanto ao **pasting ensemble** ou **floresta aleatória** ?

R: Sim, ele pode acelerar o treinamento de conjuntos ensacados e colados distribuídos por vários servidores. Todos os preditores podem ser treinados em diferentes subconjuntos aleatórios do conjunto de treinamento, permitindo, assim, várias amostras de exemplos de treinamento em vários preditores Porque cada preditor no conjunto é independente um do outro. Quanto às florestas aleatórias, você pode treinar um conjunto de classificadores de árvore de decisão em diferentes subconjuntos aleatórios do conjunto de treinamento para fazer previsões.É necessário obtê-los de todas as árvores individuais e prever a categoria com mais votos.

4. Qual é o benefício da avaliação **out-of-bag** ?

R: Na avaliação fora da bolsa, exemplos não treinados são usados para avaliar cada preditor no conjunto de bolsas. Portanto, podemos avaliar o próprio conjunto calculando a média das avaliações prontas de cada preditor sem a necessidade de um conjunto de verificação separado ou de validação cruzada. Por fim, forneceremos mais exemplos de treinamento, seu dispositivo pode melhorar um pouco o desempenho.

5. que torna as **árvores-extras** (**e xtra-trees**) mais aleatórias do que as **florestas aleatórias** comuns? Como essa aleatoriedade extra pode ajudar? As **árvores-extras** são mais lentas ou mais rápidas que as **florestas aleatórias** comuns?

R: A razão pela qual as árvores extras são mais aleatórias do que as florestas aleatórias comuns é que limiares alternativos são usados para cada recurso, em vez de encontrar o limite ideal. Essa aleatoriedade adicional pode ajudar alterando o viés para uma variação menor, que pode treinar árvores adicionais mais rapidamente, porque encontrar na árvore o melhor limite possível para cada recurso em cada nó. Uma das tarefas que requer mais tempo no desenvolvimento. As árvores em excesso são muito mais rápidas que as florestas aleatórias comuns, porque não buscam o limiar ideal; no entanto, elas não serão mais rápidas ou mais lentas que as florestas aleatórias ao fazer previsões.