

Exercício 1: Cite cinco diferenças entre o AdaBoost e o GBM?

- a. A primeira diferença é que no AdaBoost todas as árvores são do tipo STUMP, ou seja, uma árvore de profundidade tamanho 1 e com duas folhas. No GBM as árvores são convencionais, ou seja, podem ter profundidades diferentes;
- b. A segunda diferença é em relação à geração das árvores. No AdaBoost o primeiro valor é derivado de um STUMP, enquanto no GBM, o primeiro valor é a média de do “Y”;
- c. A terceira diferença é que os resultados das aplicações do algoritmo AdaBoost tem pesos diferentes para cada STUMP. No GBM além dos pesos serem diferentes, cada peso é multiplicado por um valor que é um multiplicador comum – Learning Rate. Eles são conhecidos como ETA e tem um único valor. Este valor deve ter até ser de até 0,10 para que o algoritmo tenha um bom desempenho;
- a. A quarta diferença é que no AdaBoost, os modelos são combinados por uma média ponderada das previsões, com pesos atribuídos aos classificadores com base em seu desempenho (modelos com menos erro recebem mais peso); No GBM, os modelos são somados de forma sequencial, onde a cada novo modelo é ajustado de modo a corrigir os erros do modelo anterior. As previsões são feitas pela soma ponderada das previsões dos modelos base.;
- b. A quinta diferença é em relação ao tratamento de erros. O AdaBoost pode ser muito sensível a outliers, pois os exemplos mal classificados (incluindo outliers) recebem um peso maior nas iterações subsequentes, o que pode levar a overfitting. O GBM, ao minimizar uma função de perda via gradiente, tende a ser mais robusto em relação a outliers, pois a perda não aumenta tão drasticamente para erros individuais em comparação com o AdaBoost.

Exercício 3: Cite cinco Hyper paramentos importantes no GBM

- a. **n\_estimators**: Número de árvores a serem construídas.
- b. **learning\_rate**: Taxa de aprendizado, que controla a contribuição de cada árvore para a previsão final.
- c. **max\_depth**: A profundidade máxima das árvores individuais.
- d. **subsample**: Fração de amostras usadas para treinar cada árvore (ajuda a evitar overfitting).
- e. **min\_samples\_split**: Número mínimo de amostras necessárias para dividir um nó.**max\_depth** (quando usando árvores de decisão como base\_estimator): Profundidade máxima das árvores de decisão. Limitar a profundidade pode ajudar a evitar overfitting, mantendo os classificadores fracos mais simples.

Exercício 5: Acessando o artigo de Jerome Friedman e pensando no nome dado no Scholastic GBM. Qual é a maior diferença entre os dois algoritmos?

**1. Uso dos Dados:**

- **GBM:** Usa **todos os dados** em cada árvore.
- **Stochastic GBM:** Usa uma **amostra aleatória** dos dados em cada árvore.

**2. Overfitting:**

- **GBM:** Tem mais chance de sofrer com **overfitting** (quando o modelo fica bom demais nos dados de treino, mas não generaliza bem para dados novos).
- **Stochastic GBM:** A **aleatoriedade** ajuda a evitar o overfitting, já que o modelo não fica tão fixado nos dados de treino.

**3. Velocidade:**

- **GBM:** Pode ser mais **demorado** porque usa todos os dados em cada iteração.
- **Stochastic GBM:** Como treina cada árvore com um **subconjunto** de dados, pode ser **mais rápido**.