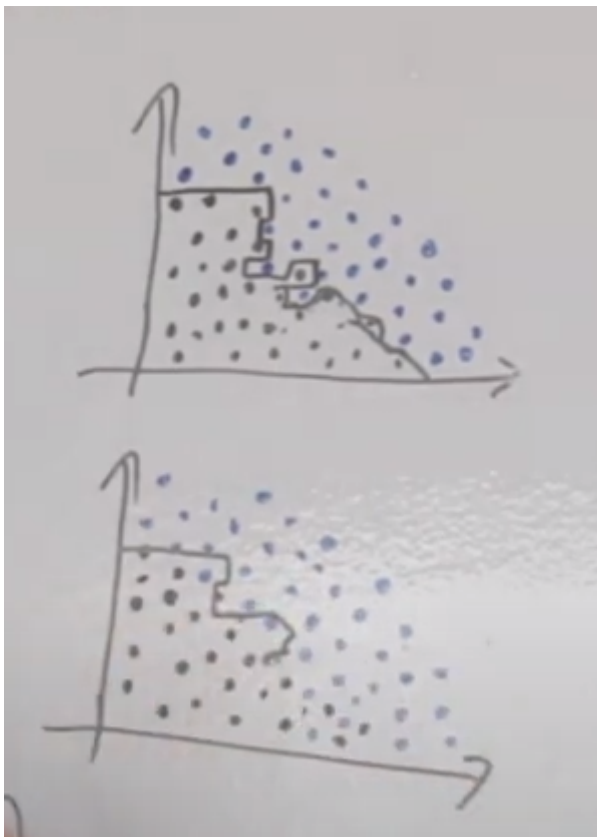


Evitando overfitting em Árvores de Decisão

Quanto mais profunda a árvore de decisão fica, menos amostras ela vai dividindo até chegar no ponto de ter somente uma amostra, entretanto podemos ter um problema se deixarmos a árvore crescer até o fim, o que chamamos de overfitting, pois a árvore vai estar perfeita para a base de treino, e não conseguirá performar bem na base de testes por ter decorado a base de treino.

Podemos observar nos gráficos abaixo que quando deixamos a árvore separar todos os dados perfeitamente criando várias retas, ela se adapta muito bem ao treino, porém nos dados de teste pode acabar errando muito mais por não serem idênticos:



Limitando o número máximo de folhas e profundidade, a árvore irá criar condições mais simples, ou seja retas mais genéricas que se adaptariam melhor no cenário desconhecido na base de testes.