



Overfitting e underfitting

## Conceito de overfitting

O termo overfitting diz respeito ao excesso de treinamento da estrutura da árvore, com isso, ela fica extremamente especializada nos dados utilizados para o treinamento e sem a capacidade de generalizar.

A capacidade de generalizar é fundamental para as árvores de decisão, pois ela diz respeito à capacidade da árvore de classificar amostras que nunca teve contato de forma correta. Um dos principais objetivos no uso das árvores é antecipar resultados de novos fatos, então não há interesse que a árvore seja excessivamente específica.

## Evitando o overfitting

Ao perceber que a árvore está ficando específica demais, é fundamental que seja feito o que se chama de poda da árvore, para reduzir a sua especificidade no conjunto de dados de treinamento e permitir que ela volte a aprender de uma forma mais genérica.

## Evitando o overfitting

O overfitting também pode ser evitado através da determinação de um tamanho máximo para a árvore, fazendo com que ela nunca cresça demais ao ponto de se tornar extremamente específica.

## Conceito de underfitting

O underfitting é o comportamento oposto ao overfitting, onde a árvore tem treinamento de menos e sequer consegue responder adequadamente aos problemas propostos.

Quando amostras do problema são apresentadas, a árvore não consegue classificá-la corretamente.

## Conceito de underfitting

Parâmetros que podem ser ajustados para controlar o nível de treinamento da árvore:

- número de amostras mínimo para divisão de um nó folha;
- número de amostras mínimo para um nó de término;
- profundidade máxima da árvore;
- número máximo de nós de término;
- número máximo de atributos para divisão.



Obrigada!

Ana Laurentino