

Mineração de Dados

Lista 3

Rafael Izbicki

Lista em trios.

NÃO COPIE! JUSTIFIQUE TODOS OS EXERCÍCIOS!

Exercício 1. Considere o banco `Carseats` disponível no pacote `ISLR` do R. O objetivo é prever a variável `Sales` (unidades vendidas) de cadeiras de bebês para carros com base nas demais.

- (a) Transforme as variáveis categóricas em dummies
- (b) Divida o conjunto de dados em 60% para treinamento e 40% para teste.
- (c) Utilizando o conjunto de treinamento, ajuste uma regressão (i) via KNN (usando validação-cruzada no treinamento para escolher os tuning parameters), (ii) via NW (usando validação-cruzada no treinamento para escolher os tuning parameters), (iii) via uma floresta aleatória e (iv) via uma árvore de regressão. Qual o melhor valor dos tuning parameters encontrados? Qual kernel você utilizou? Mostre a árvore obtida.
- (d) Qual dos métodos acima apresentou melhores resultados? Como esses resultados se comparam aos obtidos na lista anterior? Faça também um IC para os riscos estimados do melhor modelo encontrado para KNN e NW.
- (e) Repita os itens (c) e (d) usando k -fold cross validation (e, quando necessário, fazendo validação-cruzada dentro de cada fold para escolher tuning parameters). Qual modelo teve melhor desempenho? Não é necessário fazer intervalos de confiança para os riscos neste item.
- (f) Interprete os resultados do KNN e da floresta aleatória com um PDP