

Lista 2 - Aprendizado de Máquina

Thais Rodrigues

20 de dezembro de 2024

1. Se você treinou 5 modelos diferentes com exatamente os mesmos dados de treino e todos alcançaram 95% de precisão, existe alguma forma de combiná-los para obter melhores resultados? Explique.
2. É possível acelerar o treinamento de um agrupamento com o método bagging? E boosting? Explique.
3. Se o seu agrupamento do Adaboost se subajustar aos dados de treino, quais hiperparâmetros você deve ajustar e como?
4. Explique como o bagging e boosting impactam o viés e a variância do estimador.
5. Ao conjunto de dados MNIST, ajuste uma floresta aleatória e compare o desempenho do modelo com o do exercício 8 da Lista 1.
6. Qual algoritmo de treinamento de regressão linear podemos utilizar para ajustar um conjunto de dados com milhares de covariáveis e 1 milhão de observações?
7. É importante escalonar as covariáveis para aplicação de regressão com regularização? Explique.
8. Suponha que você está ajustando um modelo utilizando o método do gradiente descendente em batch e plote o erro de validação em cada época. Se o erro de validação aumenta constantemente, o que pode estar acontecendo? Como corrigir?
9. Suponha que você esteja usando a regressão Ridge e nota que o erro do treino e o erro da validação são parecidos e altos. Então o modelo estimado tem um viés alto ou variância alta? Você deve aumentar o parâmetro α da regularização ou reduzi-lo?
10. Qual é a principal vantagem da regressão LASSO em relação a regressão Ridge?
11. Ajuste uma regressão polinomial com regularização ao banco de dados Boston ($y = \text{medv}$, $x = \text{lstat}$).