

Trabalho 1

Regressão Linear

Universidade Federal do Ceará
Aprendizado de Máquina
Prof. Victor Farias



Entrega: 07/08/2020

Entrega Código + Relatório via Sigaa

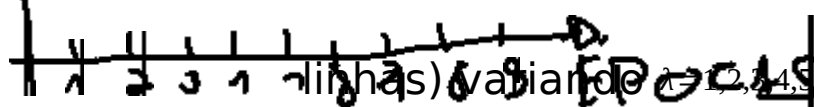
A atividade permite o uso das bibliotecas numpy e matplotlib. Mas não a scikit-learn. Os métodos devem ser implementados usando a interface mostrada em sala de aula (métodos fit e predict)

Execute as seguintes tarefas de implementação e comente o que se pede

1. Implemente os seguintes métodos
 - a. Regressão Linear univariada - método analítico
 - b. Regressão Linear univariada - gradiente descendente
 - c. Regressão Linear multivariada - método analítico (não esquecer de adicionar termo de bias)
 - d. Regressão Linear multivariada - gradiente descendente
 - e. Regressão Linear multivariada - gradiente descendente estocástico
 - f. Regressão quadrática usando regressão múltipla
 - g. Regressão cúbica usando regressão múltipla
 - h. Regressão Linear Regularizada multivariada - gradiente descendente

2. Implemente as funções
 - a. $MSE(y_{true}, y_{predict})$
 - b. $R^2(y_{true}, y_{predict})$
3. Carregue o conjunto de dados **Boston House Price Dataset** (<https://archive.ics.uci.edu/ml/machine-learning-databases/housing/>). Nesse link também contém a descrição dos atributos
4. Vamos analisar apenas a variável **LSTAT** como atributo preditor e a variável **MEDV** como atributo alvo
5. Embaralhe as amostras com seus valores alvo. Divida o conjunto de dados em 80% para treino e 20% para teste.
6. Para cada um dos métodos **a**, **b**, **f** e **g** da questão 1 faça o seguinte:
 - a. Reporte MSE e R^2 score para o conjunto de treino e o de teste
 - b. Reporte os coeficientes
 - c. Comentar qual ficou melhor a partir das métricas de erro. Descrever a razão.
7. Agora vamos analisar um segundo conjunto de dados. Carregue o conjunto de dados `trab1_data.csv` (o vetor alvo é a última coluna)
8. Para cada um dos métodos **c**, **d**, **e** e **h** (com $\lambda=1$) da questão 1 faça o seguinte:
 - a. Reporte MSE e R^2 score para o conjunto de treino e o de teste
 - b. Reporte os coeficientes
 - c. Apenas para o método **d** e **e**, plote o MSE para cada época em um gráfico linha. Comente qual dos métodos converge mais rápido.
 - d. Apenas para o método **h**, plote o MSE para o conjunto de treino e o conjunto de teste (duas

459
21



linhas) variando $\lambda = 1, 2, 3, 4, 5$. Comente qual seria o valor de λ mais adequado