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
   1. Regressão Linear univariada - método analítico
   2. *Regressão Linear univariada - gradiente descendente*
   3. **Regressão Linear multivariada – método analítico (não esquecer de adicionar termo de bias)**
   4. ***Regressão Linear multivariada – gradiente descendente***
   5. ***Regressão Linear multivariada – gradiente descendente estocástico***
   6. Regressão quadrática usando regressão múltipla
   7. Regressão cúbica usando regressão múltipla
   8. ***Regressão Linear Regularizada multivariada – gradiente descendente***
2. Implemente as funções
   1. MSE(y\_true, y\_predict)
   2. R2(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:
   1. Reporte MSE e R2 score para o conjunto de treino e o de teste
   2. Reporte os coeficientes
   3. 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 ) da questão 1 faça o seguinte:
   1. Reporte MSE e R2 score para o conjunto de treino e o de teste
   2. Reporte os coeficientes
   3. Apenas para o método **d** (Regressão Linear multivariada – gradiente descendente )e **e** (Regressão Linear multivariada – gradiente descendente estocástico),plote o MSE para cada época em um gráfico linha. Comente qual dos métodos converge mais rápido.
   4. Apenas para o método **h**, plote o MSE para o conjunto de treino e o conjunto de teste (duas linhas) variando . Comente qual seria o valor de mais adequado

