

Universidade Federal do Ceará
Campus Sobral
Engenharia da Computação e Engenharia Elétrica

Tópicos Especiais em Telecomunicações I (ECO0080)
(Reconhecimento de Padrões)

Regressão e Clustering

0) Informações Gerais

- Trabalho Individual ou de Dupla. Simulação (código) e resposta às perguntas teóricas.
- O código deve estar bem organizado e comentado, para que possa estar inteligível.
- Não usar “funções prontas” para: Regressão múltipla, Ridge Regression.
- Enviar as respostas e o código/implementação (Matlab / Python...) para o email:

david.coelho@sobral.ufc.br

- Prazo para entrega: 31/08/2021 às 23:59.

1) Base de dados:

- 1.1) Predição do preço de uma casa utilizando regressão.

<https://www.kaggle.com/harlfoxem/housesalesprediction>

- 1.2) Predição da curva de potência de um aerogerador.

“aerogerador.dat”

2) Questões:

2.1) Regressão Linear Múltipla

l) Utilizando o banco de dados “kc_house_data.csv”, encontre um modelo que descreva a relação linear entre o preço de uma casa (variável da coluna 3) com as demais variáveis (variáveis preditivas – da coluna 4 à 21).

- Utilize a regressão de cumeeira (Ridge Regression), com constante de regularização (λ) igual a 0.01.

- Utilize validação cruzada com 5 subconjuntos (5-fold cross-validation).

- Para cada “fold”, indicar o modelo obtido e o seu respectivo coeficiente de terminação ajustado R_{aj}^2 .

II) O modelo estimado de regressão (linear) “se ajustou” bem aos dados? Justifique.

2.2) Regressão Polinomial

I) De posse do conjunto de N pares entrada-saída (vento, potência), assuma que a curva de regressão é um polinômio de ordem k ($k > 0$), ou seja:

$$p(v) = a_0 + a_1*v^1 + a_2*v^2 + \dots + a_k*v^k$$

em que v é a velocidade do vento (m/s) e p é a potência gerada (kW) predita pelo modelo polinomial. Para valores de k distintos (1, 2, 3, 4 e 5). faça:

- Dividir os dados entre treinamento (70%) e teste (30%)
- Utilize a regressão de cumeieira (Ridge Regression), com constante de regularização (λ) igual a 0.01.
- Plote a curva do modelo gerado, juntamente com os dados.
- Calcule o coeficiente de determinação ajustado R_{aj}^2 .

II) Qual seria o melhor valor de k para a aproximação desta curva de geração de potência? Justifique.