

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

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

PCA e Classificador SVM

0) Informações Gerais

- Trabalho Individual. 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: SVM e PCA.
- Pode usar “funções prontas” para cálculo de: Determinante, Inversa, Autovetores e Autovalores, bem como funções utilizadas para calcular a solução do problema de programação quadrática (Ex: quadprog, fmincom).
- Enviar as respostas e o código/implementação (Matlab / Python...) para o email:

david.coelho@sobral.ufc.br

- Prazo para entrega: 05/10/20 às 23:59.

1) Base de dados: Base de classificação a sua escolha (com atributos já calculados).

- Pode ser retirado de sites, tais como:

<https://archive.ics.uci.edu/ml/datasets>

<https://www.kaggle.com/datasets>

- Restrição: A base de dados deve ter um número elevado de atributos (≥ 30), de modo que o PCA possa ser útil.

2) Questões:

2.1) Apresentação da base de dados.

- Definir o problema abordado pela base de dados.
- Evidenciar de quantas amostras, atributos e classes o problema é composto.

2.2) PCA

- I) Subtrair a média de cada atributo da base de dados.
- II) Aplicar a técnica PCA e indicar, pela “variância explicada”, a quantidade de atributos que vão ser utilizados na base de dados transformada.;
- III) Transformar a base de dados a partir da matriz de transformação gerada pela técnica PCA.

2.3) SVM

I) Informações gerais

- Implementar o SVM em sua forma dual, não-linear e relaxada.
- Sua rotina deve testar diferentes valores da constante de relaxamento “C”.
- Os seguintes kernels devem ser utilizados: Linear e RBF.
- Para o kernel RBF, testar diferentes valores para o parâmetro “gamma”.
- No caso de uma base de dados com mais de duas classes, utilizar a abordagem 1 vs all.

II) Com a base de dados transformada:

- Dividir os dados entre treinamento (70%) e teste (30%)
- Para cada valor dos hiperparâmetros “gamma” (no caso do kernel RBF) e “C”, realizar o seguinte procedimento:
 - A partir dos dados de treinamento, gere o modelo do(s) classificador(es) SVM.
 - Classifique os dados de teste (a partir do modelo gerado), e guarde a acurácia.
- A partir do modelo que gerou a melhor acurácia:
 - Indique os hiperparâmetros que levaram a esse resultado.
 - Gere a matriz de confusão obtida através dos resultados com os dados de teste.