

Exercícios de Aprendizado de Máquina

Data de Entrega: 02/12/2019

Não utilize funções prontas (algoritmos aprendidos em sala de aula). Implemente as suas e apresente-as na lista. Faça um relatório explicando como foram resolvidos os exercícios e envie junto com o código fonte para patrick.ciarelli@ufes.br

Parte V – Métodos de Agrupamento

- 1) Para as bases de dados Spiral e Jain (disponíveis em <http://cs.joensuu.fi/sipu/datasets/>), agrupe os dados em 3 e 2 grupos, respectivamente, usando kmeans e clusterização hierárquica. Avalie o resultado com a métrica de acurácia com o seguinte procedimento: para cada cluster verifique qual foi a classe predominante, amostras pertencentes a outras classes estão no grupo errado. Faça os experimentos com a distância Euclidiana. Gere gráficos com os grupos formados pelo kmeans e clusterização hierárquica. Comente os resultados. Lembre-se de não usar o atributo da classe para agrupar os dados.
- 2) Use o Apriori algoritmo para encontrar as regras de associação do conjunto de transações no arquivo transacoes.txt. Considere que os valores iguais a 1 indique a presença do item na transação. Para referência, a primeira coluna é x1, a segunda é x2 e assim em diante. Estabeleça os valores de confiança mínima e suporte de forma a encontrar algumas regras de associação. Apresente as regras encontradas.

Parte VI – Métodos Biologicamente Inspirados

- 3) Treine uma rede de Hopfield com as letras “B” e “U” contidas na pasta letras. Teste a rede com as letras “L” e “P”. Mostre e comente os resultados. Lembre-se de colocar os valores binários em -1 e 1.
Para ler as imagens no Matlab use o comando: `data = double(imread('B.bmp'));`
Para converter para vetor: `data = data(:);`
Para converter em matriz e visualizar a saída: `imshow(reshape(saida,9,7))`
- 4) Para a base de dados Concrete Compressive Strength (disponível em <http://archive.ics.uci.edu/ml/>), embaralhe os dados. Use 75% dos dados para treinar uma GRNN e 25% para teste. Use o conjunto de treino para selecionar o melhor sigma. Calcule a medida de acurácia sobre o conjunto de teste. Repita esse procedimento 4 vezes e retorne a acurácia obtida em cada teste. Evite usar sigma menor do que 0,01.
- 5) Para a base de dados Iris (disponível em <http://archive.ics.uci.edu/ml/>), embaralhe os dados. Use 75% dos dados para treinamento e 25% para teste. Use as mesmas divisões para treinar uma MLP e uma ELM. Calcule a medida de acurácia sobre o conjunto de

teste e os tempos de treinamento. Repita esse procedimento 4 vezes e retorne a média e o desvio padrão das acurácias obtidas nos testes e dos tempos de treinamento. Comente os resultados.

Questões Teóricas

- 1) Informe como pode ser usado o resultado do algoritmo DBSCAN para clusterizar novos dados, ou seja, uma vez que foi aplicado o algoritmo para encontrar os clusters, como ele é aplicado para clusterizar novos dados?
- 2) Considere um problema de classificação em que há duas classes (com 1000 amostras de treino cada, e cada amostra possui 4 atributos). Como poderia ser usado MMG na rede neural PNN para reduzir o tamanho da rede? Considere que seja usado MMG de duas componentes (com matriz de covariância completa), compare a quantidade de parâmetros que seriam armazenados na PNN clássica e na PNN com MMG. Comente a análise.