
Lista de Exercícios - Inteligência Artificial (PCS3438)

Prof. Eduardo R. Hruschka

Observações:

- Pode-se usar softwares (e.g., R, Python, Weka etc.) para resolver as questões.
- Todos os arquivos CSV possuem cabeçalho com o nome das colunas e campos separados por vírgula “,”.
- Todas as questões têm pesos iguais e valem 2,5 pontos.
- Data máxima para entrega: 30/11 no horário da aula.

Questão 1) Considerando os dados presentes no arquivo `class01.csv`, treine o algoritmo *Naive Bayes* Gaussiano utilizando a metodologia de validação cruzada *holdout* (utilize para treino as 350 primeiras linhas e para validação as demais). Qual o valor da acurácia a base de treino? Qual o valor da acurácia a base de validação?

Questão 2) Considerando os dados presentes no arquivo `class02.csv`, treine o algoritmo *10-Nearest Neighbors* (*KNN* com $k = 10$ e distância Euclidiana), utilizando a metodologia de validação cruzada *k-fold* com 5 *folds* não estratificados. Considere que a primeira pasta de validação seja formada pelas primeiras 20% linhas do arquivo, que a segunda pasta de validação seja formada pelas 20% linhas seguintes, e assim por diante, até atingir a última pasta, formada pelas 20% linhas finais da base. Qual o valor médio da acurácia para a base de validação?

Questão 3) Considerando os dados presentes no arquivo `reg01.csv`, obtenha um modelo de regressão linear com regularização L1 (*LASSO* com $\alpha = 1$) utilizando a metodologia *Leave-One-out*. Qual o valor médio do *Root Mean Squared Error* (*RMSE*) para a base de treino e para a base de validação?

Questão 4) Considerando os dados presentes no arquivo `reg02.csv`, treine uma árvore de regressão (sem realizar podas) com quebras baseadas no erro quadrático médio (do inglês *MSE* - *Mean Squared Error*) utilizando a metodologia de validação cruzada *k-fold* com $k = 5$. Qual o valor do *Mean Absolute Error* (*MAE*) para a base de treino? Qual o valor médio do *MAE* para a base de validação?