

Exercícios sobre Pré-processamento de Dados

Dica: em casos considere o tipo de atributo para realizar o pré-processamento: numérico, ordinal, periódico etc.

1. (10 pontos) Converta todos os atributos da base “Student Performance” para numéricos.
`https://archive.ics.uci.edu/ml/datasets/Student+Performance`
2. (10 pontos) Converta todos os atributos da base “Forest Fires” para numéricos.
`https://archive.ics.uci.edu/ml/datasets/Forest+Fires`
3. (15 pontos) Converta a base “Car Evaluation” de modo que esta possa ser utilizada em um experimento do tipo Holdout 50/50 estratificado com o classificador 1-NN utilizando distância Euclidiana. Calcule a taxa de acerto do classificador.
`https://archive.ics.uci.edu/ml/datasets/Car+Evaluation`
4. (15 pontos) Converta os atributos numéricos para atributos categóricos da base “Iris”. Então realize um experimento do tipo Holdout 50/50 estratificado para calcular a taxa de acerto na base transformada utilizando o classificador Árvore de Decisão.
`https://archive.ics.uci.edu/ml/datasets/Iris`
5. (15 pontos) A base “Adult” possui alguns valores de atributos omissos. Realize o experimento descrito abaixo utilizando o classificador Árvore de Decisão.
`https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data`
`https://archive.ics.uci.edu/ml/datasets/Adult`
 - (a) Divida a base em treino (50%) e teste de forma estratificada. Determine como preencher os valores omissos dos atributos utilizando apenas o conjunto de treinamento.
 - (b) Preencha os valores omissos no conjunto de treino.
 - (c) Preencha os valores omissos no conjunto de teste utilizando o método e os valores definidos para o conjunto de treino.
 - (d) Calcule a taxa de acerto do classificador.
6. (15 pontos) Faça o mesmo da questão anterior para a base “Heart Disease (hungarian)” utilizando o classificador 1-NN com distância Euclidiana.
`https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.hungarian.data`
`https://archive.ics.uci.edu/ml/datasets/Heart+Disease`

Dica: considere remover completamente uma coluna quando o valor deste característica está omissa para a maior parte dos exemplos do conjunto de treinamento.
7. (20 pontos) Treine uma rede neural MLP por 500 épocas. Este rede tem apenas uma camada escondida 10 neurônios. Avalie a taxa de acerto na base Wine com 70% dos dados de cada classe para treino e o restante para teste em dois casos distintos:
 - com os casos sem pré-processamento;
 - com cada atributo ajustado para o intervalo (0,1).