

Machine Learning

N1

Prof. Claudinei Dias (Ney)

Observações:

1. Esta avaliação contempla os seguintes assuntos: Introdução à *Machine Learning* (ML); Aprendizagem Supervisionada (paradigmas: memorização, simbólico, conexionistas e estatístico); Algoritmos classificadores (KNN, J48, ANN, NB, SVM, etc.).
2. O trabalho deve ser realizado em grupos com no mínimo 3 e no máximo 5 integrantes. Exceções devem ser solicitadas antecipadamente.

Introdução:

O Reconhecimento de Padrões, representa a capacidade de tomada de decisão com base na classe atribuída a um objeto, a um ser vivo ou a qualquer coisa que faça parte de nossa vida cotidiana. Seguem alguns exemplos: reconhecimento de caracteres, reconhecimento facial, identificação de objetos, identificação de placas e sinais, etc. Segundo Duda et al. [1], um sistema de reconhecimento de padrões permite atribuir classes a cada amostra de uma base de dados considerando as características intrínsecas ao conjunto de dados.

Considerando que a tarefa de classificação, basicamente, é dividida em duas etapas (treinamento e teste), o objetivo é projetar um programa classificador com aprendizagem supervisionada e empregando ao menos 2 (dois) classificadores em um *Data Set* presente no repositório *Machine Learning Repository* [2] ou outro *Data Set* que possua rótulos verdadeiro da classe. Uma vez que, na etapa de treinamento supervisionado o classificador aprende a distribuição dos dados através de amostras disponíveis com respectivos rótulos da classe, para posteriormente ser utilizado na inferência de novas amostras durante a etapa de teste.

Segundo Kuncheva [3], é importante saber a performance do classificador, para isto emprega-se o componente mais importante: a acurácia de classificação para obter a estimativa de precisão. Neste contexto, como é difícil experimentar todas as entradas possíveis, deve ser apresentado um relatório com a estimativa da precisão exibindo a acurácia (taxa de acerto) de cada rodada de teste realizado. Além da acurácia, outras métricas consideradas também, são: taxa de erro de classificação; a acurácia média; desvio padrão; matriz de confusão (com as métricas Precisão e Sensibilidade/Recall); um gráfico do tipo Curva ROC (*Receiver Operating Characteristics*) para exibir a comparação entre os classificadores.

Referências

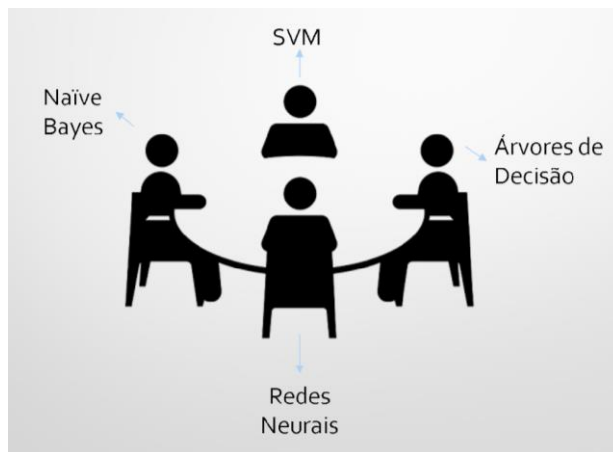
[1] Duda, Richard O.; Hart, Peter E.; Stork, David G. Pattern Classification, 2001.

[2] UC Irvine Machine Learning Repository, <http://archive.ics.uci.edu/ml/index.php>, maio de 2018.

Roteiro de atividades:

Objetivo:

Criar um programa classificador (comitê de Classificadores) com aprendizagem supervisionada e empregando ao menos 2 (dois) classificadores em um *Data Set* que possua rótulos verdadeiros da classe. Deve escolher um *Data Set* para aplicar a classificação. Após a classificação, deverá comparar os classificadores para chegar à conclusão de qual algoritmo de aprendizado de máquinas teve o melhor resultado.



Comitê de Classificadores

Algoritmos:

- KNN
- Árvores de Decisão
- Redes Neurais
- Naive Bayes
- SVM
- etc.

Etapas para documentação:

- Pesquisar e escolher um *Data Set*. **Estudar** o *Data Set* para conhecer as instâncias, as classes, as características dos atributos e as quantidades de instâncias e atributos, entre outras informações pertinentes.
- Definir** as informações do *Data Set*, das Instâncias e dos Atributos. Deve descrever a Base de Dados, explicando as variáveis, o que cada uma representa, quais os tipos de variáveis.
- Definir** qual será a variável *target* (rótulo da classe) que usará no projeto. Variável *target* será a variável de saída, a variável que se deseja “prever” algo, qual será a conclusão que se quer chegar.
- Escolher** no mínimo 2 (dois) algoritmos de *Machine Learning* para criar o comitê de classificadores.
- O documento deve conter os resultados obtidos em cada Classificador. Apresentar as seguintes **métricas**: acurácia (taxa de acerto); taxa de erro; matriz de confusão com as métricas Precisão e Sensibilidade/Recall, F1. Esses resultados devem estar justificados.
- Comparar** os resultados do Comitê de Classificadores. Usar um gráfico do tipo Curva ROC (*Receiver Operating Characteristics*) para exibir a comparação entre os classificadores.
- A **entrega** deve ser feita por um link **Google Colab** (outros formatos devem ser solicitados a priori).