

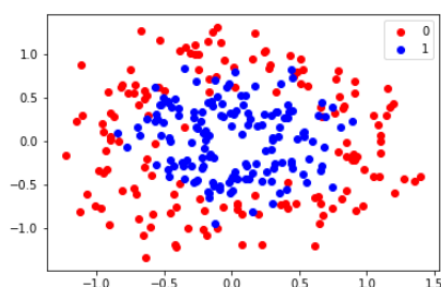
Mineração Estatística de Dados:

Projeto 3: Seleção e transformação de atributos, avaliação de classificadores

Entrega: O projeto deve ser entregue na forma de um Notebook do Jupyter no formato html.

Enviar para: **projetosicmc@gmail.com**

1 - Considere a base dados gerado pela função `skdata.make_circles` (`noise = 0.2`), conforme feito em aula.



Compare os seguintes algoritmos: (a) Redes Neurais, (b) Naive Bayes, (c) Knn, (d) Random forest. Use as métricas Accuracy, F1, Precision, Recall. Use o método k-fold cross validation com $k = 10$.

Considere a partir de agora a base de dados sobre diabetes:
<https://www.kaggle.com/uciml/pima-indians-diabetes-database>

2 - Verifique como a classificação muda quando consideramos o método k-fold cross validation. Ou seja, faça um gráfico da medida Precision em função de k para os algoritmos Redes Neurais e Random Forest.

3 - Compare os classificadores Redes Neurais, Naïve Bayes e Random Forest para $k=10$ (k-fold cross validation estratificado) em termos da curva ROC. Qual classificador é o melhor? Justifique.

4 - Usando o método que seleciona apenas um atributo por vez (univariado, teste qui-quadrado), faça uma curva da medida F1 em função do número de atributos considerados. Isto é, obtenha F1 para o atributo mais importante, depois para os dois mais importantes e assim por diante.

5 - Projete os dados usando PCA e faça um gráfico de F1 em termos do número de components. Quantos components oferecem a maior taxa de acerto? Compare com a curva “cumulative explained variance” versus “number of components” e explique os resultados.

6 – Considere agora a base de dados do Titanic, conforme descrita em:
<https://paulovasconcellos.com.br/competicao-kaggle-titanic-tutorial-5b11993774f7>

Construa a curva ROC para as redes neurais e random forest.
Identifique os atributos mais importantes usando o método wrapper. Construa um gráfico da medida Precision em função do número de atributos usados. É possível melhorar a classificação considerando um número reduzido de atributos?