

Avaliação Somativa 2

Aluno: Marcio Vinicius de Souza da Rocha

A) Treinamento e avaliação de um modelo para problema de classificação de câncer de mama a partir de imagem digital, em maligno ou benigno. A base foi criada por pesquisadores da Universidade de Wisconsin nos Estados Unidos.

Link para a base no Scikit Learn:

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html

Leitura do arquivo:

```
from sklearn.datasets import load_breast_cancer  
X, y = load_breast_cancer(return_X_y=True)
```

Qtde de atributos igual a 31, sendo:

classe (0=maligno ou 1=benigno)

30 atributos contínuos extraídos do núcleo das células em imagem de câncer de mama.

Tipo dos atributos: numéricos

Qtde de instâncias: 569 (212 malignos e 357 benignos)

Protocolo experimental a ser utilizado: validação cruzada com **5 folds**.

Encontre uma solução avaliando técnicas monolíticas e *ensembles*.

- Alternativas de ensembles de classificadores a serem avaliadas: `BaggingClassifier()` e `RandomForestClassifier()`
- Alternativas de classificadores monolíticos a serem avaliados: `MLPClassifier()` e `SVC()`.

Apresente abaixo (na tabela) o melhor resultado observado e abaixo da tabela insira a matriz de confusão.

Busque pela melhor acurácia (taxa de acerto) e o menor esforço computacional.

Apenas o Melhor Resultado

Técnica	Parâmetros	Taxa de Acerto (%)
SVC	<code>(C=1, gamma='scale', kernel='rbf', probability=True, random_state=10)</code>	97,36%

- < insira aqui a matriz de confusão>

```
[[ 81  4]
 [ 2 140]]
```

- Responda: qual a taxa de acerto de cada classe?

Classe 0 (malignos): 95,29%

Classe 1 (benignos): 98,59%

```
scaler = StandardScaler().fit(X)
X = scaler.transform(X)
X_train, X_val, y_train, y_val = train_test_split(X, y,
train_size=0.4, random_state=42, stratify=y)
svm.fit(X_val, y_val)
```

B) Treinamento e avaliação de modelo de regressão para estimativa do nível de progressão de diabetes de uma pessoa.

Nome do arquivo: diabetes

Protocolo experimental a ser utilizado: validação cruzada com 5 folds.

Link para a base no Scikit Learn

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_diabetes.html

Qtde de atributos igual a 10, sendo:

10 atributos numéricos (idade, sexo, massa corpórea, média pressão sangue, dentre outros)

Alvo = nível progressão diabetes (valor entre 25 e 346)

Tipo dos atributos: numéricos

Qtde de instâncias: 442

```
X_train, X_val, y_train, y_val = train_test_split(X, y,  
train_size=0.4, random_state=10)
```

```
svm.fit(X_val, y_val)
```

Tabela de Resultados

Técnica	Parâmetros	MAE (Mean Absolute Error)
SVR	<pre>(C=100, epsilon=0.5, gamma='scale', kernel='rbf')</pre>	44,63%

- Alternativas de ensembles de regressores a serem avaliados: BaggingRegressor() e RandomForestRegressor().
- Alternativas de regressores monolíticos a serem avaliados: MLPRegressor() e SVR().

ENTREGA EM ATÉ 48h.

HORÁRIO LIMITE: 24h de 16/06/2022

Forma de entrega: PDF no AVA.

Anexar documento no formato PDF !!!!!