

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
INSTITUTO METROPOLE DIGITAL
IMD1101 - Aprendizado de Máquina – 2022.2
Prática de Comitês de Classificadores

Visando exercitar os conceitos aprendidos nas aulas de comitês de classificadores, o aluno deve implementar 4 métodos de comitês de classificadores, Bagging, Boosting, Random Forest e Stacking.

BAGGING

Para esta estrutura de comitês de classificadores, utilize a técnica Bagging através da implementação **sklearn.ensemble.BaggingClassifier** para montar duas tabelas, uma para Bagging padrão e uma para Bagging com feature selection, no seguinte formato:

Estratégia	10	20	Media(TAM)
AD			
k-NN			
NB			
MLP			
Media(Class)			

Para isso, para a primeira tabela (Bagging padrão), faça os seguintes passos:

1. Fixar **base_estimator** para **DecisionTreeClassifier()**;
2. Variar o atributo **n_estimators** para 10 (default) e 20;
3. Para os demais parâmetros **max_samples** e **max_features**, deixe-os no seus valores default;
4. Preencher a primeira linha da tabela. Repetir os passos 1 a 3, variando o parâmetro **base_estimator** para **KNeighborsClassifier()**, **GaussianNB()** e **MLPClassifier()**, para preencher as linhas 2, 3 e 4, respectivamente.

Para a segunda Tabela (Bagging com feature selection), modifique o atributo **max_features = 0.5** e repita os passos 1 a 4.

Compare as duas tabelas, analisando seus resultados e responda a seguinte pergunta: Qual o impacto da seleção de atributos no Bagging?

BOOSTING

Para esta estrutura de comitês de classificadores, utilize a técnica Boosting através da implementação **sklearn.ensemble.AdaBoostClassifier** para montar apenas UMA tabela, no seguinte formato:

Estratégia	10	20	Media(TAM)
AD			
NB			
Media(Class)			

Para isso, para esta tabela, faça os seguintes passos:

1. Fixar **base_estimator** para **DecisionTreeClassifier()**;

2. Variar o atributo **n_estimators** para 10 (default) e 20;
3. Para os demais parâmetros, deixe-os no seus valores default;
4. Preencher a primeira linha da tabela. Repetir os passos 1 a 3, variando o parâmetro **base_estimator** para **GaussianNB()** para preencher a segunda linha 2 da tabela.

Compare esta tabela com as linhas 1 e 3 do Bagging padrão e responda a seguinte pergunta: Quem forneceu a maior acurácia, Bagging ou Boosting?

RANDOM FOREST

Para esta estrutura de comitês de classificadores, utilize a técnica Random Forest através da implementação **sklearn.ensemble.RandomForestClassifier** para montar apenas UMA tabela, no seguinte formato:

Metricas	10	100	Media(TAM)
Gini			
Entropy			
Log-loss			
Media(Med)			

Para isso, para esta tabela, faça os seguintes passos:

1. Fixar **criterion** para **gini**;
2. Variar o atributo **n_estimators** para 10 e 100 (default);
3. Para os demais parâmetros, deixe-os no seus valores default;
4. Preencher a primeira linha da tabela. Repetir os passos 1 a 3, variando o parâmetro **criterion** para **entropy** e **log_loss**, para preencher as linhas 2 e 3, respectivamente.

Compare esta tabela com os resultados do Bagging e Boosting e responda a seguinte pergunta: Qual comitê de classificadores está fornecendo a melhor acurácia? Explique sua resposta.

STACKING

Para esta estrutura de comitês de classificadores, utilize a técnica Stacking através da implementação **sklearn.ensemble.StackingClassifier**. Para criar a tabela do Stacking heterogêneo, utilize tamanhos com 10 e 20 classificadores bases. Para os algoritmos de classificação, utilize apenas MLP e k-NN, evitando configurações complexas do MLP (que demandem muito tempo).

Estratégia	Media
10 classificadores	
20 classificadores	

Compare esta tabela com os resultados do Bagging, Boosting e Random Forest e responda a seguinte pergunta: Qual comitê de classificadores está fornecendo a melhor acurácia? Explique sua resposta.

Observe que, caso haja dúvidas quanto a execução dos comitês de classificadores, recorra ao material da Aula26 postado no SIGAA.

Bom trabalho!