

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
INSTITUTO METROPOLE DIGITAL
IMD1101 - Aprendizado de Máquina – 2024.2
Prática de Comitês de Classificadores

Visando exercitar os conceitos aprendidos nas aulas de comitês de classificadores, o aluno deve implementar 5 métodos de comitês de classificadores, *Bagging*, *Boosting*, *Random Forest*, *Voting* e *Stacking*.

BAGGING

Para essa 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	30	Media (Acc TAM)
AD				
k-NN				
NB				
MLP				
Media (Acc Classificadores)				

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

1. Fixar **estimator** para **DecisionTreeClassifier**(melhor configuração);
2. Variar o atributo **n_estimators** para 10 (default), 20 e 30;
3. Para os demais parâmetros **max_samples** e **max_features**, deixe-os com seus valores default;
4. Preencher a primeira linha da tabela. Repetir os passos 1 a 3, variando o parâmetro **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. Por último, 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	30	Media (Acc TAM)
AD				
NB				
Media (Acc Classificadores)				

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

1. Fixar **estimator** para **DecisionTreeClassifier**(melhor configuração);
2. Variar o atributo **n_estimators** para 10 (default), 20 e 30;
3. Para os demais parâmetros, deixe-os com seus valores default;

4. Preencher a primeira linha da tabela. Repetir os passos 1 a 3, variando o parâmetro **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:

Estratégia	10	20	30	100	Media (Acc TAM)
Gini					
Entropy					
Log-loss					
Media (Acc Classificadores)					

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

1. Fixar **criterion** para **gini**;
2. Variar o atributo **n_estimators** para 10, 20, 30 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 5, 10, 15 e 20 classificadores bases. Para os algoritmos de classificação, utilize MLP, k-NN, DT e NB, evitando configurações complexas do MLP (que demandem muito tempo).

Estratégia	Media
5 classificadores	
10 classificadores	
15 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.

VOTING

Para esta estrutura de comitês de classificadores, utilize a técnica *Voting* através da implementação **sklearn.ensemble.VotingClassifier**. Para criar a tabela do *Voting*

heterogêneo, utilize tamanhos com 5, 10, 15 e 20 classificadores bases. Para os algoritmos de classificação, utilize MLP, k-NN, DT e NB, evitando configurações complexas do MLP (que demandem muito tempo).

Estratégia	Media
5 classificadores	
10 classificadores	
15 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 Aula24 postado no SIGAA.

Bom trabalho!