

Nama: Krisnhadi Bima

NIM: 1301164014

Ensemble Learning

Ensemble learning adalah sebuah teknik yang digunakan pada machine learning untuk meningkatkan hasil akurasi prediksi dengan menggabungkan lebih dari 1 model. Ensemble learning bertujuan untuk mengurangi variansi (*bagging*), bias (*boosting*), dan meningkatkan akurasi (*stacking*).

Bagging

Bagging atau **B**ootstrap **A**ggregation, adalah teknik untuk mengurangi variansi pada dataset. Setiap model pada Ensemble Learning, melakukan training pada subset dataset (dipilih secara random).

Untuk menghasilkan output, setiap model melakukan klasifikasi terhadap data uji, kemudian dengan melakukan voting pada setiap model, dapat ditarik kesimpulan terhadap hasil akhir klasifikasi.

Tugas Program

Untuk tugas program, terdapat 270 data training. Untuk implementasi Ensemble Learning, saya menggunakan NaiveBayes (sesuai dengan soal), dan tambahan model SVM dan Regression.

```
pydev.py pydev_executor.py debugger.py ptvsd_launcher.py naive_bayes.py train.csv ensemble.py X
C:\Program Files (x86)\Microsoft Visual Studio\Shared\Python36_64\python.exe
pydev debugger: starting
C:\Program Files (x86)\Microsoft Visual Studio\Shared\Python36_64\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default solver
r will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)
C:\Program Files (x86)\Microsoft Visual Studio\Shared\Python36_64\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The default value of gamma
will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this w
arning.
"avoid this warning.", FutureWarning)
Train OK
{'data': array([[34.1, 5. ]]), 'votes': [1, 1, 1], 'result': 1, 'correct': True}
{'data': array([[14.05, 21.4 ]]), 'votes': [2, 2, 1], 'result': 2, 'correct': True}
{'data': array([[17.8, 14.2 ]]), 'votes': [1, 1, 1], 'result': 1, 'correct': True}
{'data': array([[38.75, 16.85 ]]), 'votes': [2, 1, 1], 'result': 1, 'correct': True}
{'data': array([[18. , 22.3 ]]), 'votes': [2, 2, 1], 'result': 2, 'correct': True}
{'data': array([[13.7, 27.75 ]]), 'votes': [1, 2, 1], 'result': 1, 'correct': False}
{'data': array([[26. , 4.25 ]]), 'votes': [1, 1, 1], 'result': 1, 'correct': True}
{'data': array([[19.4, 8.85 ]]), 'votes': [1, 1, 1], 'result': 1, 'correct': True}
{'data': array([[39.25, 15.5 ]]), 'votes': [1, 1, 1], 'result': 1, 'correct': True}
{'data': array([[26.2, 4.4 ]]), 'votes': [1, 1, 1], 'result': 1, 'correct': True}
nv 8 out of 10 0.8
reg 10 out of 10 1.0
svm 7 out of 10 0.7
overall, 9 out of 10
class of [ 3.3 15.45 2. ] is 2
class of [ 3.9 21.85 2. ] is 2
class of [ 4.6 18.25 2. ] is 2
class of [ 5.2 23.4 2. ] is 2
class of [ 7.05 19.9 2. ] is 2
class of [ 9.75 25.5 1. ] is 1
class of [11.65 26.85 1. ] is 1
class of [12.75 22.05 2. ] is 2
class of [14.05 26.55 1. ] is 1
class of [14.15 17.35 2. ] is 2
class of [14.2 22.15 2. ] is 2
class of [15.2 11.7 1. ] is 1
class of [15.95 16.45 2. ] is 2
class of [16.05 11.9 1. ] is 1
class of [16.5 15.15 1. ] is 1
class of [16.95 9.7 1. ] is 1
class of [17.1 14. 1. ] is 1
class of [17.25 13.15 1. ] is 1
class of [17.3 14.9 1. ] is 1
```

Gambar 1: Hasil ensemble learning dengan menggunakan 10 data uji saja (data uji disampling secara random dari 270 data training)

Program dapat dirun di file ensemble.py

Library sklearn diperlukan untuk program ini.