Ensemble methods: boosting

By Marios Michailidis

Examined ensemble methods

- Averaging (or blending)
- Weighted averaging
- Conditional averaging
- Bagging
- Boosting
- Stacking
- StackNet

What is Boosting

A form of weighted averaging of models where each model is built sequentially via taking into account the past model performance.

Main boosting types

- Weight based
- Residual based

Weight based boosting

Rownum	x0	x1	x2	x3	y
0	0.94	0.27	0.80	0.34	1
1	0.84	0.79	0.89	0.05	1
2	0.83	0.11	0.23	0.42	1
3	0.74	0.26	0.03	0.41	0
4	0.08	0.29	0.76	0.37	0
5	0.71	0.76	0.43	0.95	1
6	0.08	0.72	0.97	0.04	0

Weight based boosting

Rownum	x0	x1	x2	x3	y	pred
0	0.94	0.27	0.80	0.34	1	0.80
1	0.84	0.79	0.89	0.05	1	0.75
2	0.83	0.11	0.23	0.42	1	0.65
3	0.74	0.26	0.03	0.41	0	0.40
4	0.08	0.29	0.76	0.37	0	0.55
5	0.71	0.76	0.43	0.95	1	0.34
6	0.08	0.72	0.97	0.04	0	0.02

Weight based boosting

Rownum	x0	x1	x2	x3	y	pred	abs.error
0	0.94	0.27	0.80	0.34	1	0.80	0.20
1	0.84	0.79	0.89	0.05	1	0.75	0.25
2	0.83	0.11	0.23	0.42	1	0.65	0.35
3	0.74	0.26	0.03	0.41	0	0.40	0.40
4	0.08	0.29	0.76	0.37	0	0.55	0.55
5	0.71	0.76	0.43	0.95	1	0.34	0.66
6	0.08	0.72	0.97	0.04	0	0.02	0.02

Weight based boosting

Rownum	x0	x1	x2	x3	y	pred	abs.error	weight
0	0.94	0.27	0.80	0.34	1	0.80	0.20	1.20
1	0.84	0.79	0.89	0.05	1	0.75	0.25	1.25
2	0.83	0.11	0.23	0.42	1	0.65	0.35	1.35
3	0.74	0.26	0.03	0.41	0	0.40	0.40	1.40
4	0.08	0.29	0.76	0.37	0	0.55	0.55	1.55
5	0.71	0.76	0.43	0.95	1	0.34	0.66	1.66
6	0.08	0.72	0.97	0.04	0	0.02	0.02	1.02

Weight based boosting

Rownum	x0	x1	x2	x3	y	weight
0	0.94	0.27	0.80	0.34	1	1.20
1	0.84	0.79	0.89	0.05	1	1.25
2	0.83	0.11	0.23	0.42	1	1.35
3	0.74	0.26	0.03	0.41	0	1.40
4	0.08	0.29	0.76	0.37	0	1.55
5	0.71	0.76	0.43	0.95	1	1.66
6	0.08	0.72	0.97	0.04	0	1.02

Weight based boosting parameters

- Learning rate (or shrinkage or eta)
- Number of estimators more estimators we add to these type of ensemble, the smaller learning rate we need to put.This is sometimes quite difficult to find the right values, and we do it with the help of cross-validation
- Input model – can be anything that accepts weights
- Sub boosting type:
  - AdaBoost – Good implementation in sklearn (python)
  - LogitBoost - Good implementation in Weka (Java)

Residual based boosting

I believe that in any predictive modeling competition that was not image classification or predicting videos. This has been the most dominant type of algorithm that actually has one most in these challenges so this type of boosting has been extremely successful

Rownum	x0	x1	x2	x3	y
0	0.94	0.27	0.80	0.34	1
1	0.84	0.79	0.89	0.05	1
2	0.83	0.11	0.23	0.42	1
3	0.74	0.26	0.03	0.41	0
4	0.08	0.29	0.76	0.37	0
5	0.71	0.76	0.43	0.95	1
6	0.08	0.72	0.97	0.04	0

Residual based boosting

Rownum	x0	x1	x2	x3	y	pred
0	0.94	0.27	0.80	0.34	1	0.80
1	0.84	0.79	0.89	0.05	1	0.75
2	0.83	0.11	0.23	0.42	1	0.65
3	0.74	0.26	0.03	0.41	0	0.40
4	0.08	0.29	0.76	0.37	0	0.55
5	0.71	0.76	0.43	0.95	1	0.34
6	0.08	0.72	0.97	0.04	0	0.02

Residual based boosting

Rownum	x0	x1	x2	x3	y	pred	error
0	0.94	0.27	0.80	0.34	1	0.80	0.20
1	0.84	0.79	0.89	0.05	1	0.75	0.25
2	0.83	0.11	0.23	0.42	1	0.65	0.35
3	0.74	0.26	0.03	0.41	0	0.40	-0.40
4	0.08	0.29	0.76	0.37	0	0.55	-0.55
5	0.71	0.76	0.43	0.95	1	0.34	0.66
6	0.08	0.72	0.97	0.04	0	0.02	-0.02

Residual based boosting

Rownum	x0	x1	x2	x3	y
0	0.94	0.27	0.80	0.34	0.2
1	0.84	0.79	0.89	0.05	0.25
2	0.83	0.11	0.23	0.42	0.35
3	0.74	0.26	0.03	0.41	-0.4
4	0.08	0.29	0.76	0.37	-0.55
5	0.71	0.76	0.43	0.95	0.66
6	0.08	0.72	0.97	0.04	-0.02

Residual based boosting

Rownum	x0	x1	x2	x3	y	new pred
0	0.94	0.27	0.80	0.34	0.2	0.15
1	0.84	0.79	0.89	0.05	0.25	0.20
2	0.83	0.11	0.23	0.42	0.35	0.40
3	0.74	0.26	0.03	0.41	-0.4	-0.30
4	0.08	0.29	0.76	0.37	-0.55	-0.20
5	0.71	0.76	0.43	0.95	0.66	0.24
6	0.08	0.72	0.97	0.04	-0.02	-0.01

Residual based boosting

Rownum	x0	x1	x2	x3	y	new pred	old pred
0	0.94	0.27	0.80	0.34	0.2	0.15	0.80
1	0.84	0.79	0.89	0.05	0.25	0.20	0.75
2	0.83	0.11	0.23	0.42	0.35	0.40	0.65
3	0.74	0.26	0.03	0.41	-0.4	-0.30	0.40
4	0.08	0.29	0.76	0.37	-0.55	-0.20	0.55
5	0.71	0.76	0.43	0.95	0.66	0.24	0.34
6	0.08	0.72	0.97	0.04	-0.02	-0.01	0.02

Residual based boosting

Rownum	x0	x1	x2	x3	y	new pred	old pred
0	0.94	0.27	0.80	0.34	0.2	0.15	0.80
1	0.84	0.79	0.89	0.05	0.25	0.20	0.75
2	0.83	0.11	0.23	0.42	0.35	0.40	0.65
3	0.74	0.26	0.03	0.41	-0.4	-0.30	0.40
4	0.08	0.29	0.76	0.37	-0.55	-0.20	0.55
5	0.71	0.76	0.43	0.95	0.66	0.24	0.34
6	0.08	0.72	0.97	0.04	-0.02	-0.01	0.02

To predict Rownum=1 we would say :

Final prediction = 0.75 + 0.20 = **0.95**

Residual based boosting

Rownum	x0	x1	x2	x3	y	new pred	old pred
0	0.94	0.27	0.80	0.34	0.2	0.15	0.80
1	0.84	0.79	0.89	0.05	0.25	0.20	0.75
2	0.83	0.11	0.23	0.42	0.35	0.40	0.65
3	0.74	0.26	0.03	0.41	-0.4	-0.30	0.40
4	0.08	0.29	0.76	0.37	-0.55	-0.20	0.55
5	0.71	0.76	0.43	0.95	0.66	0.24	0.34
6	0.08	0.72	0.97	0.04	-0.02	-0.01	0.02

To predict Rownum=1 we would say :

Final prediction = 0.75 + 0.20 = **0.95**

Residual based boosting parameters

- Learning rate (or shrinkage or eta)
- Number of estimators
- Row (sub) sampling
- Column (sub) sampling
- Input model – better be trees.
- Sub boosting type:
  - Fully gradient based
  - Dart

Dart, it imposes a drop out mechanism in order to control the contribution of the trees. This is a concept derived from deep learning where you say, "Every time I make a new prediction in my sample, every time I add a new estimate or I'm not relying on all previous estimators but only on a subset of them."

Residual based favourite implementations

- Xgboost
- Lightgbm
- H2O's GBM
- Catboost
- Sklearn's GBM