MachineLearnAthon - Microlecture LightGBM

Recorded by Lara Kuhlmann

MachineLearnAthon
A project Co-funded by the Erasmus+ programme of the European Union

Learning outcomes of today

After successfully completing this micro-lecture, you are able to….

- Understand the advantages of LightGBM, including its speed and memory efficiency
- Describe the purpose of a loss function in machine learning and how it is used to quantify prediction differences
- Grasp how Exclusive Feature Bundling (EFB) reduces features by identifying and merging exclusive features using Greedy Bundling and Merge Exclusive Features algorithms
- Comprehend how Gradient-based One-Side Sampling (GOSS) reduces data instances while maintaining accuracy by focusing on samples with larger gradients
- Learn how the histogram-based algorithm efficiently finds optimal split points in decision trees
- Understand the overall process of composing decision trees in LightGBM

Co-funded by the Erasmus+ programme of the European Union

Agenda for today

- Advantages
- Main components
- Loss function
- Exclusive Feature Bundling (EFB)
- Greedy Bundling
- Merge Exclusive Features
- Gradient-based One-Side Sampling (GOSS)
- Histogram-based algorithm
- Composing decision trees

Co-funded by the Erasmus+ programme of the European Union

Advantages

- Training process of the gradient boosting decision tree (GBDT) is up to over 20 times faster with LightGBM with almost the same accuracy in the predictions[1]
- Especially with large amounts of data[1,2]
- But even with low-dimensional data, LightGBM is faster than XGBoost[2]
- Uses less ram[2,3]

[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.
[2] Fu, F.; Jiang, J.; Shao, Y.; Cui, B. (2019): An Experimental Evaluation of Large Scale GBDT Systems.
[3] Gouri, M. H. & Kumar, M. (2023): Empowering Machine Learning with Scikit-Learn, XGBoost and LightGBM.

Main components

- Reduction of features with Exclusive Feature Bundling (EFB)[1]
- Reduction of the data points with Gradient-based One-Side Sampling (GOSS)[1]
- Splitfinding with histogram-based algorithm[1]
  - ➢ Combination of trees

[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.

Loss function

- Central concept in machine learning.
- Quantification of the difference between the predictions & the actual values. [8,9]

➢ The smaller the value, the more accurate the prediction.

- Goal: To minimize the value.
- Quadratic loss function[8,9] : $L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$

[8] Hastie, T.; Tibshirani, R.; Friedman, J. (2009): Boosting and Additive Trees. In: The Elements of Statistical Learning. Springer, New York. S. 337-387.
[9] Friedman, J. H. (2001): Greedy function approximation: A gradient boosting machine. In: The Annals of Statistics, 5 (2001) 29. S. 1189 – 1232.

Exclusive Feature Bundling (EFB)

- This method is defined by two algorithms:
  1. Which features are bundled 🎭 Greedy Bundling[1]
  2. How the bundle is constructed 🎭 Merge Exclusive Features[1]
- High-dimensional data often very sparse.
  - ➢ It is rare for columns to contain two values at the same time(not equal to 0).[1]

[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.

Greedy Bundling

- Create a Conflict Count Matrix based on the conflicts of the individual features and sort the totals.

Daten-pkt.	F0	F1	F2	F3	F4
0	1	1	0	0	1
1	0	0	1	1	1
2	1	2	0	0	2
3	0	0	2	3	1
4	2	1	0	0	3
5	3	3	0	0	1
6	0	0	3	0	2
7	1	2	3	4	3
8	1	0	1	0	0
9	2	3	0	0	2

conflict count matrix

	F0	F1	F2	F3	F4
F0		6	2	1	6
F1	6		1	1	6
F2	2	1		3	4
F3	1	1	3		3
F4	6	6	4	3	
∑	15	14	10	8	19

searchOrder

	F4	F0	F1	F2	F3
∑	19	15	14	10	8

Bild: [1] Tabellen: [10]

[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.
[10] [MXML-12-03] Light GBM [3/5] - Exclusive Feature Bundling (EFB), Greedy Bundling. Online verfügbar unter: https://www.youtube.com/watch?v=Y-IvfsjmqOQ (zuletzt aufgerufen 13.11.2024).

Gradient-based One-Side Sampling (GOSS)

Reduce the number of data instances while maintaining the accuracy of the learned decision trees.[1]

- Gradient provides useful information for sampling data for each data point in GBDT.[1,12]
- More attention is paid to samples with larger gradients during training.[1,4,5]

[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.
[4] Yang, H.; Qin, G.; Liu, Z.; Hu, Y.; Dai, Q. (2024): LightGBM robust optimization algorithm based on topological data analysis. In: Proceedings of the 2024 International Conference on Computer and Multimedia Technology, S. 574 – 582.
[5] Zhang, D. & Gong, Y. (2020): The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure. In: IEEE, 8 (2020), S. 220990-221003.
[12] Jung, A. (2024): Gradientenbasiertes Lernen. In: Machinelles Lernen. Springer Nature, Singapore. S. 109-123

Gradient-based One-Side Sampling (GOSS)

Data point	Living space m^2	Rental price €
1	30	400
2	45	600
3	50	700
4	60	800
5	25	350
6	70	900
7	80	1000
8	90	1100
9	65	850
10	100	1200

Data point	Living space m^2	Rental price €
11	110	1300
12	55	750
13	75	950
14	85	1050
15	40	550
16	120	1400
17	95	1150
18	105	1250
19	115	1350
20	35	450

[8] Hastie, T.; Tibshirani, R.; Friedman, J. (2009): Boosting and Additive Trees. In: The Elements of Statistical Learning. Springer, New York. S. 337-387.
[9] Friedman, J. H. (2001): Greedy function approximation: A gradient boosting machine. In: The Annals of Statistics, 5 (2001) 29. S. 1189 – 1232.

Gradient-based One-Side Sampling (GOSS)

1. Sorting the data points according to the absolute value of their gradients.[1]
2. Data reduction by selecting the “upper” instances (a) and randomly selecting a part of the remaining data (b).[1]
3. Amplification of the data with small gradients by constant.[1] $\frac{(1-a)}{b}$

[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.

Gradient-based One-Side Sampling (GOSS)

1.1 Initialisation[5,9,13]

- Initial forecast (f_0): ∅ Rental price

$$f_0 = \frac{1}{n}\sum_{i=1}^{n} yi = \frac{(400+600+\ ...+1400)}{20} = 875\ €$$

[5] Zhang, D. & Gong, Y. (2020): The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure. In: IEEE, 8 (2020), S. 220990-221003.
[9] Friedman, J. H. (2001): Greedy function approximation: A gradient boosting machine. In: The Annals of Statistics, 5 (2001) 29. S. 1189 – 1232.
[13] Seto, H.; Oyama, A.; Kitora, S.; Toki, H.; Yamamoto, R.; Kotoku, J.; Haga, A.; Shinzawa, M.; Yamakawa, M.; Fukui, S.; Moriyama, T. (2022): Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data. Scientific Report, 12, Artikel: 15889.

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Gradient-based One-Side Sampling (GOSS)

1.2 Calculation of gradient[8]

- Using loss function $L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$
- Gradient g_i with regard to prediction $\hat{y}$

$$g_i = \frac{\partial L(y, \hat{y})}{\partial \hat{y}_i} = \hat{y}_i - y_i = 875 - y_i$$

[8] Hastie, T.; Tibshirani, R.; Friedman, J. (2009): Boosting and Additive Trees. In: The Elements of Statistical Learning. Springer, New York. S. 337-387.

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Gradient-based One-Side Sampling (GOSS)

Data point	Rental price €	Prediction $\hat{y}_i$	Gradient g_i
1	400	875	475
2	600	875	275
3	700	875	175
4	800	875	75
5	350	875	525
6	900	875	-25
7	1000	875	-125
8	1100	875	-225
9	850	875	25
10	1200	875	-325

Data point	Rental price €	Prediction $\hat{y}_i$	Gradient g_i
11	1300	875	-425
12	750	875	125
13	950	875	-75
14	1050	875	-175
15	550	875	325
16	1400	875	-525
17	1150	875	-275
18	1250	875	-375
19	1350	875	-475
20	450	875	425

Gradient-based One-Side Sampling (GOSS)

1. Sorting

- Sort data points by absolute value of the gradients.

Data point	Gradient $\|g_i\|$
5	525
16	525
1	475
19	475
11	425
20	425
18	375
10	325
15	325
2	275

Data point	Gradient $\|g_i\|$
17	275
8	225
3	175
14	175
7	125
12	125
4	75
13	75
6	25
9	25

Gradient-based One-Side Sampling (GOSS)

2. Data reduction

- Top 20% with the largest $|g_i|$
- ➢ 4 data points
- 30% of the remaining data points
- ➢ 6 data points

Data point	Gradient $|g_i|$
5	525
16	525
1	475
19	475
11	425
20	425
18	375
10	325
15	325
2	275

Data point	Gradient $|g_i|$
17	275
8	225
3	175
14	175
7	125
12	125
4	75
13	75
6	25
9	25

Gradient-based One-Side Sampling (GOSS)

3. Adjustment of the weightings

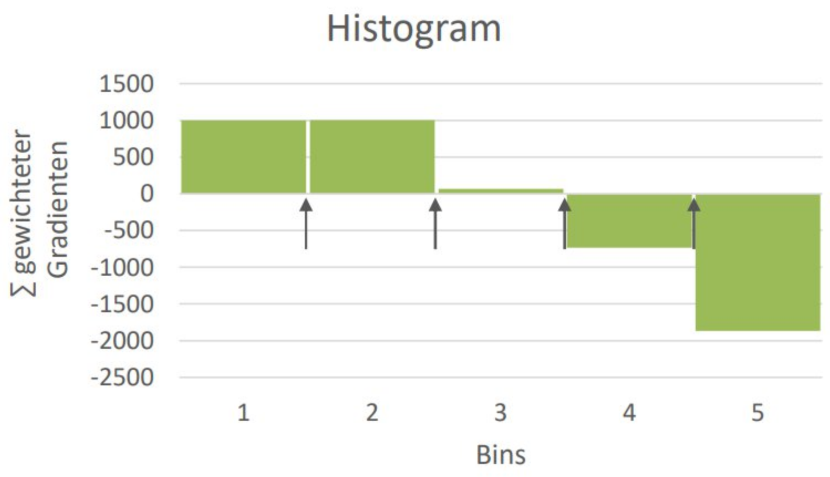
- Weighting of the fully selected “top” data points: ω_i = 1
- Weighting of randomly selected data points:

$$\omega_i = \frac{(1-a)}{b} = \frac{0{,}8}{0{,}3} = 2{,}67$$

Histogram-based algorithm

Finding the optimal split of the feature (living space), to reduce the loss function as much as possible.

1. Preparation of the data[1,2]
2. Creating the histograms[1,2]
3. Calculation of the split gains for possible splits[1,2]



[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.
[2] Fu, F.; Jiang, J.; Shao, Y.; Cui, B. (2019): An Experimental Evaluation of Large Scale GBDT Systems.

Further procedures

- Recalculation of the gradients with the new predictions[1]
- Apply the GOSS method to the child nodes[1]
- Perform split finding until termination conditions are met[1]
- As long as training is not completed
  - Tree is created in each new iteration (attempts to reduce errors from previous trees)[1,8]
- GOSS procedure is applied when selecting the data to train the new tree[1]

[1] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. (2017): LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, S. 3149 – 3157.
[8] Hastie, T.; Tibshirani, R.; Friedman, J. (2009): Boosting and Additive Trees. In: The Elements of Statistical Learning. Springer, New York. S. 337-387.

Recap this lecture

After successfully completing this lecture, you are able to….

- Understand the advantages of LightGBM, including its speed and memory efficiency
- Describe the purpose of a loss function in machine learning and how it is used to quantify prediction differences
- Grasp how Exclusive Feature Bundling (EFB) reduces features by identifying and merging exclusive features using Greedy Bundling and Merge Exclusive Features algorithms
- Comprehend how Gradient-based One-Side Sampling (GOSS) reduces data instances while maintaining accuracy by focusing on samples with larger gradients
- Learn how the histogram-based algorithm efficiently finds optimal split points in decision trees
- Understand the overall process of composing decision trees in LightGBM

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

This material was created with the assistance of an AI language model.

Co-funded by the Erasmus+ programme of the European Union