

2.Sitzung

25.06.2021

Tree-Methods

Vorteile von Decision- and Regression-Trees

1. orthogonale Splits im Variablen-Raum; keine non-linearen Trennlinien

Random-Forest

2. Wie funktioniert es
3. Vor- und Nachteile von Random-Forest
4. Viele "Tree-Stumps" zusammen können non-lineare Trenn- oder Regressionslinien approximieren

Optimization

5. Warum sind quadratische Fehlerfunktionen leichter zu optimieren als absolute Fehler
6. Kenne das Prinzip von Newton-Raphson (Optimierungs-Verfahren 1.Ordnung)
7. Wie funktioniert Stochastic Gradient Descent (SGD)
8. Maximieren der Likelihood ist universeller als die Reduktion eines quadratischen Fehlers
9. Taylor-Series ermöglichen die Approximierung von Funktionen durch die Ableitungen der Funktion an nur einem Punkt
10. Approximation der Likelihood durch eine Taylor-Series und Ableitung nach der Schrittweite führt zu einem Optimierungs-Verfahren 2. Ordnung

Gradient Boosted Trees (GBT)

11. Likelihood wird berechnet zwischen tatsächlichen Werten und augenblicklich vorhergesagten Werten
12. Einem Parameter-Update entspricht beim Gradient Boosting Trees die weitere Hinzunahme eines Tree-Predictors zur augenblicklichen Vorhersage
13. Die Likelihood bei Schritt t kann durch die likelihood zu Schritt t-1 durch eine Taylor-Series approximiert werden
14. Das Vorhersage-Update zu Schritt t ergibt sich analog zur Newton Optimierung 2.Ordnung
15. Für Gradient Boosting Regression ist das Update die mittleren Residuen aller Beobachtungen in einem Leaf-Node.
16. Gradient Boosting Regression kann also - etwas unsauber - als Regression der bisherigen Residuen bezeichnet werden

- 17. L1 und L2 Regularization bezieht sich auf die Updates, die mit jedem weiteren Baum zu Schätzung hinzugefügt werden
- 18. Wir kennen die wichtigsten Parameter ein Gradient Boosted Trees
- 19. Wir wissen woher das "Gradient" im Namen herkommt

House-Prices mit Gradient Boosting Machine (GBM)

- 20. Preprocessing-Steps wie bei Elastic-Net bewirken nichts
- 21. Interaktionen werden durch die Trees automatisch modelliert, monotone Transformationen der Variablen wirken sich nicht auf Trees aus
- 22. Grösste Verbesserungen durch Hyper-Parameter Fine-Tuning
- 23. Evaluate a "good" idea in an un-biased way, d.h. ohne overfitting

Feature-Importance

- 24. Feature-Importance als wichtiger Baustein zur Interpretation von Tree-Methods
- 25. Nachteil der Berechnung über Reduktion der Gini-Impurity
- 26. Permutation-Importance ist weniger biased (bessere Wahl)
- 27. Bei gini-Impurity hängt die Feature-Importance von der Kardinalität einer Variablen ab
- 28. Die Wichtigkeit einer Variablen kann durch One-Hot-Encoding auf viele One-Hot-Variablen aufgeteilt werden und erscheint fast als unwichtig
- 29. Redundante Variablen (Korrelation mit anderer Variable sehr hoch) können problemlos entfernt werden - obwohl sie eine hohe feature Importance haben