

2.Sitzung

25.06.2021

Tree-Methods

Vorteile von Decision- and Regression-Trees

1. Verstehen, was der Fehler einer Linearen Regression ist
2. orthogonale Splits im Variablen-Raum; keine non-linearen Trennlinien

Random-Forest

3. Wie funktioniert es
4. Vor- und Nachteile von Random-Forest
5. Viele "Tree-Stumps" zusammen können non-lineare Trenn- oder Regressionslinien approximieren
6. Polynomiale Regression als Beispiel für Multivariate Lineare Regression; Einsicht, dass die Formel unverändert bleibt, die Regressions-Linie aber keine Gerade mehr ist.

Optimization

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

Gradient Boosted Trees (GBT)

13. Likelihood wird berechnet zwischen tatsächlichen Werten und augenblicklich vorhergesagten Werten
14. Einem Parameter-Update entspricht beim Gradient Boosting Trees die weitere Hinzunahme eines Tree-Predictors zur augenblicklichen Vorhersage
15. Die Likelihood bei Schritt t kann durch die likelihood zu Schritt $t-1$ durch eine Taylor-Series approximiert werden
16. Das Vorhersage-Update zu Schritt t ergibt sich analog zur Newton Optimierung 2.Ordnung

- 17. Für Gradient Boosting Regression ist das Update die mittleren Residuen aller Beobachtungen in einem Leaf-Node.
- 18. Gradient Boosting Regression kann also - etwas unsauber - als Regression der bisherigen Residuen bezeichnet werden
- 19. L1 und L2 Regularization bezieht sich auf die Updates, die mit jedem weiteren Baum zu Schätzung hinzugefügt werden
- 20. Wir kennen die wichtigsten Parameter ein Gradient Boosted Trees
- 21. Wir wissen woher das "Gradient" im Namen herkommt

House-Prices mit Gradient Boosting Machine (GBM)

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

Feature-Importance

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