

Übungsblatt 3

1. Gradio mit Random Forest für eine Regression



Verwenden Sie Gradio, um eine GUI für eine Regression mit Random Forest für strukturierte Daten mit einer einzelnen numerischen Zielgröße ($\mathbb{R}^n \rightarrow \mathbb{R}$) zu erstellen. Sie sollen davon ausgehen, dass die Daten als csv vorliegen und mit Pandas geladen werden dürfen.

Testen Sie ihren Code an Hand des California Housing Datensatz. Dieser Datensatz enthält Informationen über Einkommen, Bevölkerung, Hausbelegung, Lage (Breitengrad, Längengrad) und allgemeine Informationen über das Haus wie Anzahl der Zimmer, Anzahl der Schlafzimmer und Alter des Hauses für jeden Bezirk in Kalifornien. Der Datensatz besteht aus 20640 Einträgen und 9 Merkmalen. Es gibt keine fehlenden Werte. Die Zielvariable ist der Median des Hauswertes für die kalifornischen Bezirke, ausgedrückt in Hunderttausend Dollar. Wir haben diese Daten in der Datei `california_housing.csv` zusammengestellt, die bereits im scibo heruntergeladen werden kann.

Weitere Informationen zum Datensatz finden Sie unter folgendem Link:
https://www.dcc.fc.up.pt/~ltorgo/Regression/cal_housing.html

Die GUI von Gradio soll folgende Elemente aufweisen:

1. Einen Schieberegler, mit dessen Hilfe, Prozentwert für die Aufteilung in Trainings- und Testmenge festgelegt werden kann.
2. Schieberegler oder numerische Eingaben zur Einstellung der Hyperparameter **Decimals**, **Threshold**, **MinLeafNodeSize** und **NoOfTrees**
3. Einen Startkopf *Trainingsstart*, der das Training startet
4. Regressionsplots, wie auf Seite 217 des Buches Maschinelles Lernen von Prof. Frochte, zur Visualisierung der Ergebnisse, inklusive einer von Ihnen designten “GUI-Mechanik“ um die Auswertung zu starten (predict).

Sie können Ihr eigenes GUI-Design erstellen, aber es sollte die oben genannten Elemente enthalten und Sie können, wenn Sie möchten, weitere Funktionalitäten hinzufügen.