

## Übungsblatt 2

### 1. Random Forest für die Klassifikation



Setzen Sie analog zu dem Random Forest für die Regression aus dem Buch nun einen Variante für die Klassifikation um. Wenn Sie Code vom Regressions-Fall übernehmen, achten Sie darauf, dass `threshold ≥ 1` keinen Sinn für die Klassifikation hat. Testen Sie Ihre Umsetzung auf dem Two Moons Problem mit 1000 Datensätzen pro Klasse. Erweitern Sie ihren Code so, das am Schluss ein Plot mit den Gebieten die einer Klasse zugeordnet werden entsteht. Gehen Sie für den Plot analog zu dem Ansatz auf Seite 76 des Buches vor und nutzen Sie `pcolormesh`.

### 2. Feature Importance, Random Forest und CART



Am Schluss des Foliensatzes zum Random Forest wird das Boston House Price Data Set untersucht. Erzeugen Sie mit ihrer Implementierung (Sie dürfen das ZIP zum Buch nutzen) des Random Forest den gleichen Plot, wie auf der letzten Folie zum Random Forest. Sie können dabei auf die schwarzen Balken für die Schwankungsbreite verzichten. Versuchen Sie mittels nur maximal vier Merkmalen eine möglichst hohe Genauigkeit unter Verwendung eines durch den CART-Algorithmus erzeugten Baumes zu erreichen. Versuche Sie den Baum dabei durch Wahl geeigneter Parameter auf weniger als 10 Blätter zu beschränken. Testen Sie die Genauigkeit auf einer Testmenge die 20% der ursprünglichen Daten umfasst.