

Applied Analytical Data Science

- Übung 4 -

Dr. Jörg-Uwe Kietz,
Übung zur Vorlesung,
Mittwoch, 14-16 Uhr Vorlesung,
16-18 Uhr Übung (alle 2 Wochen)

<http://www.kietz.ch/AADS/>

Gewicht

- $W = \text{abs}(\text{Target}_D - \text{Porto}) / \text{max_Target}_D$

Vorhanden von den letzten Übungen

- 1) Alle relevanten Daten in der nützlichsten Form fürs Data Mining
- 2) Keine fehlenden Werte mehr in den Daten
- 3) Split der Lerndaten in 66% Training und 33% Test
- 4) Modelle zur Vorhersage wer spendet
- 5) Eine CPE (Class Probability Estimation) wer spendet
- 6) Profit-Chart für 4)

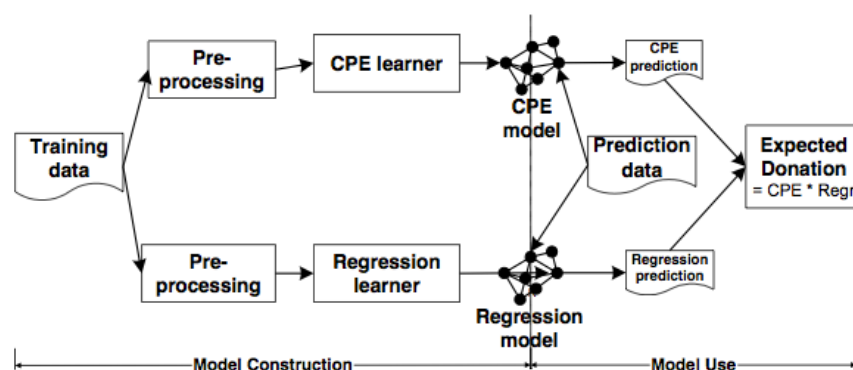
4. Aufgabe: Lernen wie viel jemand spendet

- Benutzt den Split der Lerndaten in 66% Training und 33% Test der letzten Übung
- Regressionslernen von Target_D
- Lernt ein Model auf den Trainingsdaten
- Benutzt „Apply Model“ auf den Test-Daten

Aufgabe 3

- RM produziert Predicted(Target_D)
- Benenne es um und ändere die Rolle Predicted_Target_D um es besser weiter benutzen zu können.
- Vergleiche die Profit Charts von
 - CPE der letzten Übung
 - Predicted_Target_D dieser Übung
 - Produkt von CPE und Predicted_Target_D (beide kombiniert)

Marketing-Process (Bernstein, et al. p. 27)



Regressions Lerner in RM

Edit/New Operator

Fill Dialog as shown

All Ops with a lamp are Learners

Profit Chart

Order the population by estimated likelihood

Sum up the Profit gained mailing the best X%

Y-axis: Cummulated Profit
X-axis: Percentage of Pop.

Vortrag jeder Gruppe in der nächsten Übung

- Welche Regressionsmethoden habt ihr probiert?
- Wie habt ihr sie Evaluert?
- Welche waren wie gut?
- Waren die Besten gut genug?
- Was kann man besser machen?
- Waren sie besser als CPE?
- War die kombination aus Regression und CPE besser?

Abgabe der 4. Übung

- Vortrag am 02.05.18 (Mittwoch in der Übung)
- RM Process, Modelle und Resultate (Report, Graphs, Validation) via Email bis 30.4.18 (Montag abend)
Bitte einen gezippten Folder namens GrXUebY mailen.