

8. Übungsblatt

Ksenia Klassen
ksenia.klassen@udo.edu

Dag-Björn Hering
dag.hering@udo.edu

Henning Ptaszyk
henning.ptaszyk@udo.edu

10. Januar 2017

1 Vorbereitungen

1.1 a)

Zu Beginn werden die Attribute mit der höchsten Korrelation zu dem Attribut SalePrice bestimmt. Diese werden mit dem Prozess Blatt8/Prozesse/Korrelationbestimmen.rmp bestimmt, die weiteren Aufgaben sind ebenfalls in diesem Prozess zu finden.

1.OverallQual	: 0,790
2.GrLivArea	: 0,708
3.GarageCars	: 0,640

In den Abbildungen 1-3 sind die jeweiligen Korrelation als Scatter-Plot dargestellt.

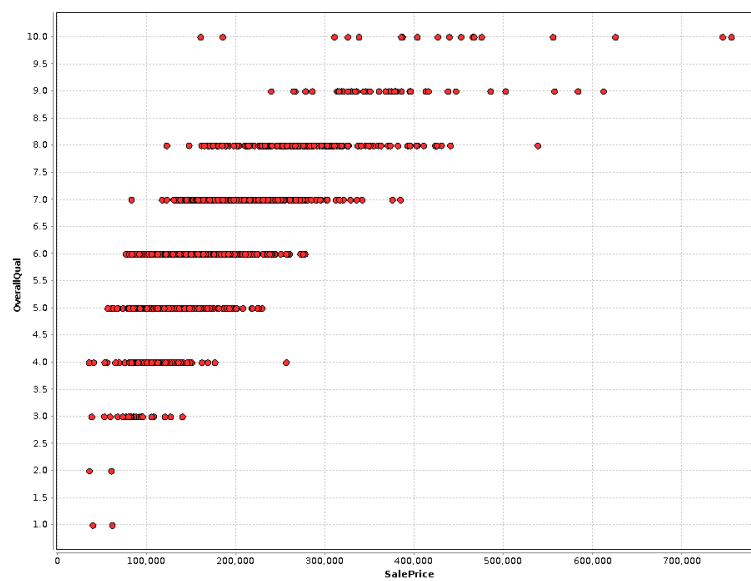


Abbildung 1: Scatter-Plot von der Korrelation von SalePrice und OverallQual.

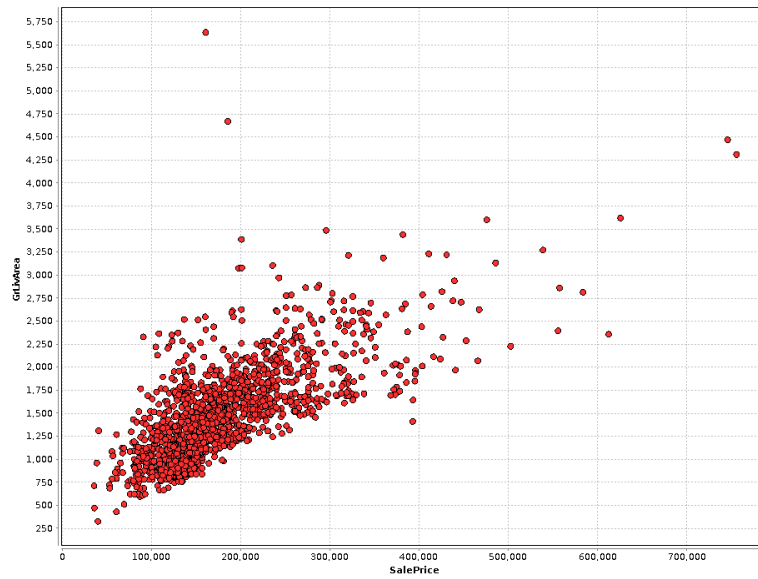


Abbildung 2: Scatter-Plot von der Korrelation von SalePrice und GrLivArea.

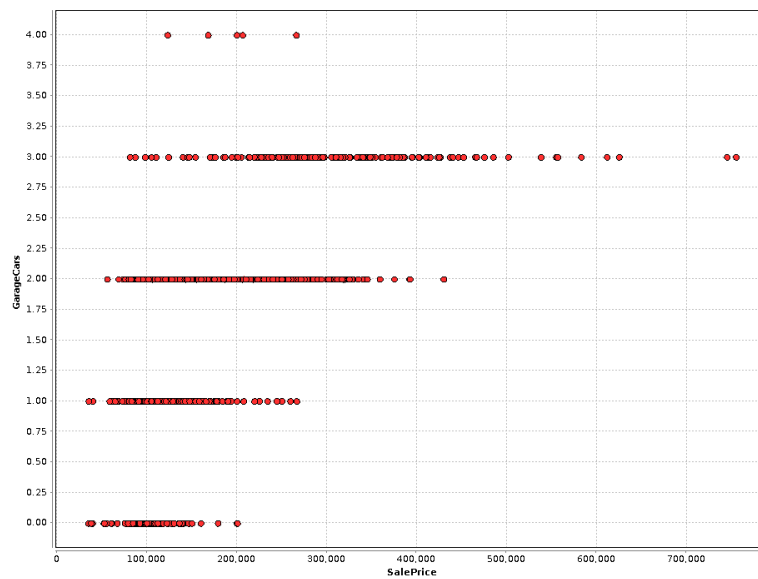


Abbildung 3: Scatter-Plot von der Korrelation von SalePrice und GarageCars.

1.2 b)

Nun wird eine lineare Regression zwischen dem Attribut OverallQual und dem Attribut SalePrice durchgeführt zu sehen in der Abbildung 4.

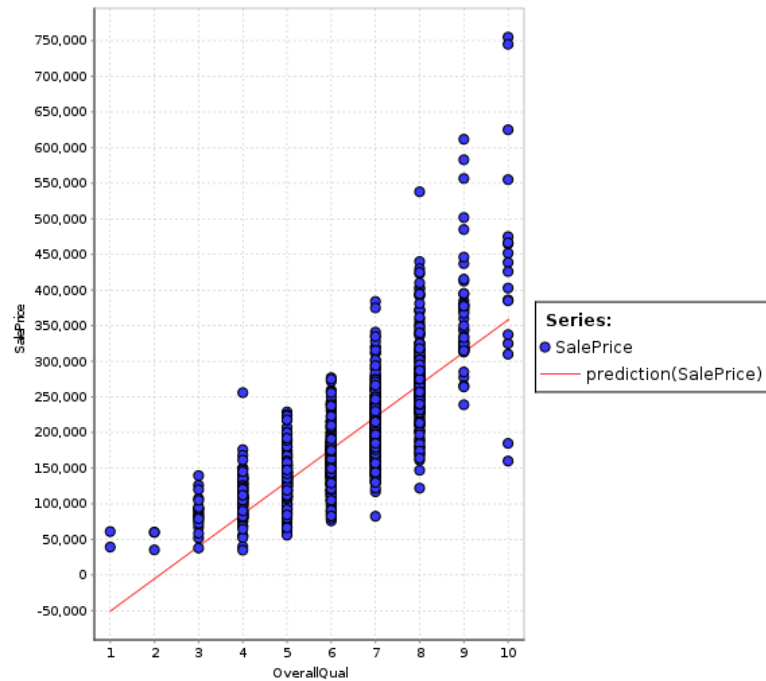


Abbildung 4: Lineare Regression zwischen SalePrice und OverallQual.

1.3 c)

In der Abbildung 5 wird der Relative Abstand zwischen dem geschätzten Verkaufspreisen aus der linearen Regression aus 1.2 und den wahren Verkaufspreisen in einem Histogramm aufgetragen.

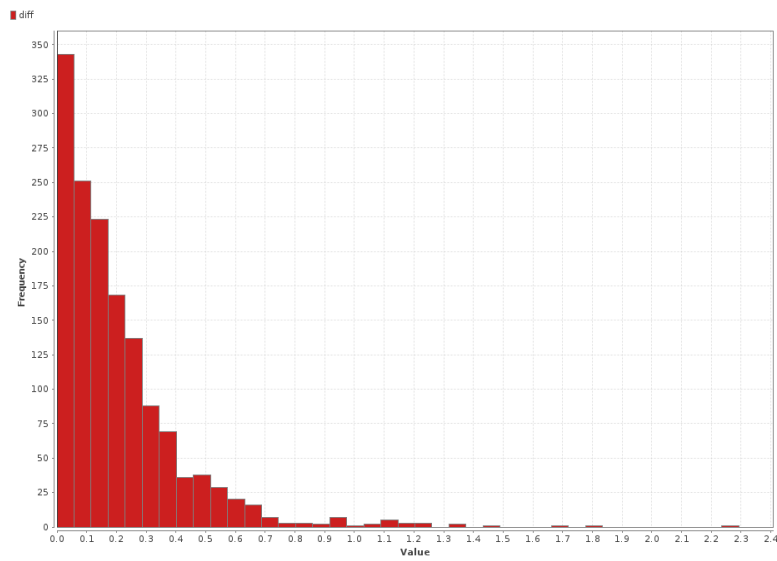


Abbildung 5: Histogramm für den Relativen Abstand von den geschätzten Verkaufspreisen zu den Wahren Verkaufspreisen.

1.4 d)

Die Aufgabe der Data-Mining-Challenge wurde mit **RapidMiner** bearbeitet. Die zugehörigen Prozesse (als Prozessdateien und Screenshots) sowie auch die **prediction.csv**, in der das Ergebnis der Regression zu finden ist, befinden sich im Ordner **MiningChallenge**.

Die Ergebnisse der Validierung werden beim Ausführen als **RESULT** ausgegeben.