

DATA MINING

- Versuch 2: Vorhersage und Clustering auf Finanzdaten -

Teammitglieder

JanHorak

Ralf Palyov

Andreas Mayer

3.1 Zeitreihenschätzung: Vorhersage des Aktienkurses

3.1.2 Kursvorhersage mit SVR

1 Überlegen Sie sich genau, wie die Datenvektoren des Vorhersagezeitraums (also die Vektoren, die der Methode prediction() des trainierten SVR-Modells übergeben werden), aufgebaut werden müssen.

2 Für welche Werte von

- Time Delay
- SVR-Parameter C
- SVR-Parameter epsilon

erreichen Sie die beste Vorhersage? Wie groß ist in diesem Fall der MAE? Erzeugen Sie für diese optimierten Werte den Plot des tatsächlichen und des geschätzten Kursverlaufs und speichern Sie diese Grafik unter dem Namen stockpredict.png.

3.2 Clustering der Aktienkursverläufe

1 Analysieren Sie die Clusterzuweisungen. Gehören die Unternehmen, welche in einem Cluster zusammenfallen, irgendwie zusammen?

Aus meiner Sicht gibt es mindestens zwei Kriterien die die Zugehörigkeit zu einem Cluster bestimmen:

- Branche
- Firmenstandort/Absatzmarkt

Viele Zuordnungen sind auf den ersten Blick einleuchtend, andere würden eine genaue Analyse der jeweiligen Unternehmen erfordern. Auf Grund der hohen Anzahl an Unternehmen (60) und Clustern (12), erwähne ich nur einige offensichtlichere Beispiele:

Im zweiten Cluster sind alle Pharma-Konzerne der Unternehmensliste zusammengefasst. Das fünfte Cluster besteht aus Banken, Kreditinstituten, Versicherungen und Mischkonzernen. Das letzte Cluster besteht aus Rüstungskonzernen und Boeing (die jedoch auch Militärflugzeuge produzieren).

Es gibt Cluster in denen relativ viele verschiedene Branchen vertreten sind, z.B. Cluster 7: Hier ist ein Zusammenhang zum Standort oder zum Hauptabsatzmarkt ersichtlich. Sony, Canon, Toyota, Honda, Mitsubishi sind japanische Konzerne, in dem Cluster finden sich allerdings auch noch zwei weitere Unternehmen Caterpillar und Unilever, bei diesen ist mir der Zusammenhang allerdings nicht klar.