



Das
Programm zur
Auswertung von
'NEAREST NEIGHBOUR'
Prediktoren

von

Dr. rer. nat. Hans Uhlig

Copyright Hinweis

Der Text und die Abbildungen dieses Beitrages sowie die zugehörige Software unterliegen dem Urheberrechtsschutz. Wer diese Produkte erwirbt, darf sie für den eigenen Gebrauch nutzen. Kopien oder Wiedergaben in anderer Form ob vollständig oder nur teilweise, bedürfen meiner schriftlichen Zustimmung.

Hans Uhlig, Oktober 2009

Auswertungsprogramme 'Nearest Neighbour' Prediktoren.

Mit Auswertungsprogrammen kann man den Wertpapier- oder Devisenhandel nach vorgegebenen Regeln simulieren und diese Simulation über einen gewissen Zeitraum laufen lassen, um zu testen, wie sich das System bewährt hätte. Ein System, das in der Vergangenheit funktioniert hat bietet zwar keine Sicherheit, dass es auch weiterhin funktionieren wird, doch es ist der einzige Anhaltspunkt, den man hat.

TradeStation^(R), MetaStock^(R) usw. taugen nur für globale Prediktoren

Institutionelle und unabhängige selbständige Börsenhändler können derzeit zwischen mehreren kommerziellen Programmen wählen, die in der Lage sind, Handelsstrategien zu prüfen und Handelsstatistiken auszugeben. An erster Stelle wäre da die TradeStation^(R) von Firma OMEGA zu nennen, aber auch andere Produkte, wie beispielsweise MetaStock^(R) von Firma EQUIS und weitere mehr bieten derartige Möglichkeiten. Diese Programme sind recht flexibel. Neben Varianten vorgegebener Strategien können sie auch neu programmierte Regeln auswerten. Es gibt jedoch und das betrifft mehrere meiner Programme, einige wesentliche Einschränkungen: erstens können sie nur globale Strategien auswerten, d.h. lokale Strategien gehören nicht zu ihrem Repertoire. Nearest Neighbour Prediktoren und Chaos-Statistik Prediktoren können also damit nicht getestet werden. Aber auch andere Formen anpassbarer Prediktoren, wie beispielsweise mein Programm PAVLOV, siehe gesonderte Beschreibung, das mit mehreren verschiedenen Strategien gleichzeitig arbeitet, können mit diesen Programmen nicht ausgewertet werden.

Globale Handelsstrategien sind solche, die bestimmte Kriterien für Marktsituationen definieren, die günstig sind für Marktengagements. Alle Daten werden geprüft, ob sie diese Kriterien erfüllen und wenn das der Fall ist, wird gehandelt. Der Erfolg der Strategie wird dadurch bestimmt, wie gut diese Kriterien die günstigen Marktbedingungen beschreiben und wie häufig solche Bedingungen vorliegen. Als Beispiel für eine einfache globale Handelsstrategie mag die von mir vorgestellte Methode des robusten Trends (Uhlig, H. 2001b) dienen.

Die Auswertungsprogramme für lokale Prediktoren

Da die allermeisten Börsenhändler globale Prediktoren verwenden, gibt es auf diesem Gebiet viel Konkurrenz und das schmälert den möglichen Gewinn. Schon aus dem Grund wäre es sinnvoll, andere Ansätze zu prüfen. So hat man eventuell die Chance, profitable Gelegenheiten zu finden, die bei herkömmlicher Betrachtung übersehen werden. Wirklich gute lokale Prediktoren werden nicht veröffentlicht, sonst wäre ihr Nutzen schnell dahin. Wer davon profitieren möchte, muss sie selbst finden, dann braucht er den Ertrag auch nicht mit anderen zu teilen. Die Suche nach geeigneten oder gar optimalen Parametern ist ohne schnelle Auswertungsprogramme eine mühsame und zeitaufwendige Angelegenheit. Mit Hilfe des hier vorgestellten Programms kann man mögliche Kandidaten für profitable Strategien schnell, ausgiebig und konsistent, d.h. widerspruchsfrei prüfen.

Ich habe mehrere Auswertungsprogramme für die 'Nearest Neighbour' Prediktoren und für die Chaos-Statistik Vorhersagen geschrieben, von denen ich das erste kurz vorstellen möchte. Wie auch für andere von mir besprochene Programme, habe ich hier wieder die wöchentlichen Schlusskurse des Deutschen Aktienindex DAX von Anfang 1993 bis Mitte 2005 verwendet. Diese Zeitreihe umfasst verschiedene Börsenphasen mit steigenden und fallenden Kursen und schien mir daher gut geeignet für Testzwecke. Damit man die Bildschirmhalte auch ohne Screenshot-Programm speichern kann, habe ich jeweils die Ausgabe in eine Datei mit vorgesehen. Falls Interesse besteht, kann ich die Programme auch für das Betriebssystem LINUX zugänglich machen.

Das Auswertungsprogramm für 'Nearest Neighbour' Prediktoren

```
MS-DOS-Eingabeaufforderung - G032
7 x 12
+++++++ direkte Nearest Neighbour Prognose ++++++
Das Programm simuliert Vorhersagen fuer 1 Zeitschritt(e)
Nun 4 Glaettungsfaktoren mit Leerzeichen getrennt eingeben:

Einbettungen einheitlich oder kombiniert:
  1 = mit 4 Zeitverzoegerungen
  2 = mit 3 Zeitverzoegerungen + 1 Differenz
  3 = mit 2 Zeitverzoegerungen + 2 Differenzen
  4 = mit 1 Zeitverzoegerung + 3 Differenzen
  5 = mit 4 Differenzen

Bitte eine Auswahl treffen (1..5): 2

Eingabe: 3 Zeitschritte und 1 Differenz (Beispiel: 1 2 4 2): 1 1 4 1
Bitte geben Sie 4 Gewichtungsfaktoren ein (Dezimalbrueche erlaubt),
mit Leerzeichen dazwischen (Beispiel: 1 1.2 1.3 1.4): 1 1 1 1

NN Auswertung waehlen (1=mit3, 2=gmit3, 3=mit4, 4=gmit4, 5=mit5, 6=gmit5): 4

Daten einlesen (max 5000), Bitte Dateinamen eingeben: dax9305.prn
Referenzdatensatz von dem begonnen werden soll !! muss > 0 sein !!: 1
Wie viele, Testdaten sollen ausgewertet werden: 640
```

oben sehen Sie den Eingabedialog für ein Auswertungsprogramm. Nicht alle Eingaben sind lesbar, denn mein Know-how möchte ich natürlich nicht verschleiern. Wichtig ist, dass man sehen kann, welche Möglichkeiten angeboten werden und wie die Eingaben erfolgen.

Die Zahl der Vorhersageschritte ist hier voreingestellt. Es ist immer ein Zeitschritt - für eine spätere Programmversion könnte ich die Zeitschritte wählbar machen, wobei zwei Zeitschritte sicherlich noch sinnvoll wären, aber jenseits von vier Zeitschritten ist keine Prognose mehr möglich, die den Namen verdient. Als nächstes wird nach vier Glättungsfaktoren gefragt. Diese beziehen sich auf vier aufeinander folgende Werte in der eingelesenen Datenreihe. Sie müssen nicht mit geglätteten Werten arbeiten. Falls Sie mit den Rohdaten arbeiten möchten, dann geben Sie einfach nur 1 0 0 0 ein. Dadurch wird der erste Wert mit 1 multipliziert und die nachfolgenden mit null. Falls Sie nur über zwei Werte glätten wollen, geben Sie 1 1 0 0 ein. Denken Sie, dass der erste Wert mehr Gewicht haben sollte, können Sie ihn auch mit 1.2 oder 1.5 oder einfach beliebig gewichten. Entsprechendes gilt für drei und mehr Gewichtungsfaktoren.

Die Einbettung kann einheitlich oder eine Kombination aus zwei Verfahren sein. Im Beispiel habe ich eine Kombination gewählt. Die ersten drei Werte sind zeitverzögert eingebettet und der vierte mit Differenz. Sie sehen, dass der erste Wert doppelt eingegeben ist, dazu der vierte und die erste Differenz.

Als nächstes kann man die Eingaben gewichten. Ich habe allen den Wert 1 gegeben. Dadurch wird der erste Wert doppelt gewichtet, weil er zwei mal auftritt. Ich hätte auch als Gewichtung 0.5 0.5 1 1 eingeben können, dann wären alle drei Werte gleich gewichtet worden.

Die anschließende Frage betrifft die Zahl der 'nächsten Nachbarn', über die das Ergebnis gemittelt werden soll : mit3, mit4 usw. sind Mittel über drei bzw. vier 'nächste Nachbarn'. Steht ein 'g' davor,

also gmit5 beispielsweise, dann bedeutet dies, dass über fünf nächste Nachbarn gemittelt werden soll. Diese sollen aber gewichtet werden und zwar umgekehrt proportional zu ihrem Abstand vom Referenzmuster. Das heißt, die ähnlichsten Werte werden am stärksten gewichtet, die unähnlichsten am geringsten. Im Beispiel wurde Modus 4 gewählt, das sind vier gewichtete 'nächste Nachbarn'.

Schließlich muss man noch die Datei eingeben, von der die Daten gelesen werden und festlegen, von welchem Datensatz begonnen und wieviele insgesamt ausgewertet werden sollen. Im Beispiel beginnt man bei 1 und lässt 640 Daten auswerten.

Sobald man die letzte Eingabe durch Drücken der Eingabetaste bestätigt hat, sieht man das Ergebnis.

# Trans.	Ertrag buy/hold	Ertrag NN.Pred.	b&h.korrekt	NN.korrekt
26	0.095	0.167	15	20
52	0.217	0.211	32	37
78	0.174	0.152	46	47
104	0.317	0.402	64	66
130	0.630	1.071	80	86
156	0.253	1.204	88	100
182	-0.063	1.214	97	112
208	-0.081	1.371	111	127
234	-0.253	1.153	124	137
260	-0.429	1.351	133	149

312	-0.109	1.560	162	180
364	-0.008	1.504	191	205
416	0.153	1.581	223	233
468	0.636	1.699	263	265
520	0.748	1.782	294	293
572	0.792	1.816	319	319
624	0.931	1.961	350	348

Ergebnisse abspeichern (1) oder nicht (0)? Bitte (1 oder 0) eingeben 0
 beliebige Taste druecken

Oben sieht man die Bildschirmausgabe des Auswertungsprogramms für die 'Nearest Neighbour' Vorhersage. In der linken Spalte ist die Anzahl der Transaktionen angegeben, d.h. die Anzahl der Wochen, für die der Test durchgeführt wurde. Die nächsten beiden Spalten zeigen die Erträge 'buy+hold', also 'kaufen und halten' und des 'Nearest Neighbour' Prediktors. Die beiden rechten Spalten geben an, wie oft die Investitionsentscheidung günstig war. Oberhalb der gestrichelten Linie sind die Erträge im Halbjahresabstand angegeben, darunter im Jahresabstand. Es werden also insgesamt die Ergebnisse von 12 Jahren ausgegeben, von Mitte 1993 bis Mitte 2005.

Das Programm zählt rückwärts vom aktuellsten Wert bis zum ältesten. Die obere Zeile ist also das Ergebnis der letzten 26 Wochen, darunter stehen die letzten 52 Wochen ... usw. Die Methode 'kaufen und halten' hat sehr schwankende Erträge, während das 'Nearest Neighbour' Programm einen nahezu kontinuierlichen Zuwachs verzeichnet. In den letzten 12 Jahren hätte es beim DAX etwa den doppelten Ertrag gebracht, jedenfalls unter den hier betrachteten Bedingungen des konstanten Investitionsvolumens. Das trifft aber bei der Methode 'kaufen und halten' nicht zu, bei der ständig die Erträge reinvestiert werden. Eine solche Strategie ist jedoch problematisch, wie ich gleich erläutern werde.

Wenn man die Erträge reinvestiert, dann verändert man auch das Anlagerisiko, denn angenommen, die Erträge sind positiv, riskiert man mit der Zeit immer höhere Verluste. In den Zeiten, als die Finanzmärkte nicht so turbulent waren, hat man das Verlustrisiko vernachlässigt. Doch die Schönwetterperiode ist vorbei. Inzwischen sollte jedem klar sein, dass man die Erträge riskanter und weniger riskanter Anlagen nicht 1 : 1 vergleichen darf. Man muss auch das Risiko berücksichtigen und das ist es, was der oben angeführte Vergleich zeigen soll, denn er bezieht sich auf ein einheitliches Risiko.

Weitere Angaben zum Programm

Anders als die meisten anderen meiner Programme, ist dieser 'Nearest Neighbour' Prediktor in der Programmiersprache 'C' geschrieben. Das Programm ist sehr rechenintensiv, wegen der vielen Sortiervorgänge und weil 'C' ist die höhere Programmiersprache, die die schnellsten Programme erzeugt, habe ich diese Sprache dafür gewählt. Auch von diesem Programm gibt es eine GUI Version, die in 'C++' geschrieben ist, weil ich so den 'C' Code vollständig übernehmen konnte. Sie sehen einen Bildschirmabdruck hier unten wiedergegeben.

'Nearest Neighbour' Prediktor als GUI-Programm

Nearest Neighbour Predictor

Datei Hilfe

Dateiangaben

Datensätze	
2007.02.23	4582.63
2007.02.16	4533.01
2007.02.09	4554.44
2007.02.02	4560.19
2007.01.26	4448.96

Dateiname: C:/hans/Hauptfenster/qfile17/atxtest2.txt Datenanzahl 18

Dateiformat

nur Zahlenwerte
 Datumangabe und Zahlenwert

Abstandsberechnung

Manhattan Metrik
 Euklidische Metrik

Einbettungsvarianten

Zeitabstände:Differenzen (4:0)
 Zeitabstände:Differenzen (3:1)
 Zeitabstände:Differenzen (2:2)
 Zeitabstände:Differenzen (1:3)
 Zeitabstände:Differenzen (0:4)

Glättung

Faktor 1: 1 Faktor 2: 0 Faktor 3: 0 Faktor 4: 0

Datenmuster

Punkt 1: 1 Punkt 2: 2 Punkt 3: 3 Punkt 4: 4

Nächste Nachbarn

4 Anzahl nächster Nachbarn

Gewichtung invers zum Abstand
 Gewichtung einheitlich

Statistikbereich

Startwert: Endwert:

Auswertung - Bericht

Das GUI-Programm entspricht funktionell ganz dem Konsolenprogramm. Auch der Auswertungsbericht wird wie bei der Konsolen-Version auf dem Bildschirm dargestellt und kann auf Wunsch als Datei abgespeichert werden.

Die GUI-Versionen meiner Programme sind in ihrer jetzigen Form jeweils nur vorläufige Entwürfe. Ich kann sie den Kundenwünschen entsprechend ändern und anpassen. Mögliche Änderungen beziehen sich nicht nur auf die äußere Form. Auf Wunsch könnte beispielsweise anstelle der bisher angegebenen halbjährlichen Zwischensummen auch eine ausführliche, detaillierte Handelsstatistik ausgegeben werden, in der jeder einzelne Vorgang aufgelistet ist.