Schriftliche Ausarbeitung zur Besonderen Lernleistung
im Fach Informatik zu dem Thema

Sortierung hochdimensionaler Daten in Excel

Name: Eric Rossa
Klasse: 12/1
Schuljahr: 2008/2009
Schule: Gymnasium Stephaneum
Betreuender Lehrer: Herr Glode
Betreuer des IPK Gatersleben: Dr. Strickert
Datum der Abgabe: 17. April 2009

Inhaltsverzeichnis

Zusammenfassung

Das von mir entwickelte Programm dient als Werkzeug zum Sortieren hochdimensionaler biologischer Genexpressionsdaten per Selbstorganisierender Merkmalskarte (SOM) als VBA-Programm in Microsoft Excel 2003. Damit ist es möglich, biologische Genexpressionsdaten aus einer Exceltabelle auszulesen, diese nach bestimmten Kriterien einzufärben und dann anschließend zeilenweise zu sortieren. Die besondere Herausforderung dabei liegt in der gleichzeitigen Verarbeitung mehrerer Spalten für die Bestimmung einer optimalen Sortierreihenfolge. Dadurch kann man dann Ähnlichkeiten verschiedener Pflanzen z.B. die von verschiedenen Gerstensorten auf der Ebene spezifisch aktivierter Gene leicht erkennen, was eine wichtige Teilaufgabe im Rahmen der Tätigkeiten der Bioinformatik-Arbeitsgruppe Dateninspektion am Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung Gatersleben ist. Die spezielle Sortierung ermöglicht es, durch ähnliche Züchtungsmerkmale die optimalen Kompositionen verschiedener Gerstensorten zu ermitteln und diese dann für Zuchtzwecke zu kombinieren. Anhand eines weiteren Beispieles mit Aminosäuremessungen wird gezeigt, dass das Programm einen allgemeinen Nutzen für die Verarbeitung biologischer Daten besitzt.

IPK Gatersleben

Das Leibnitz-Institut für Pflanzengenetik und Kulturpflanzenforschung ist eines der bedeutensten Forschungseinrichtungen für die Pflanzenforschung.
Vorrangig wird im IPK mit Kulturpflanzen gearbeitet [1].
Im Zentrum grundlagen- und anwendungsorientierten Forschung steht das Erringen neuer Erkenntnisse und Technologien mit dem Ziel einer umfassenden Nutzung pflanzengenetischer Ressourcen für eine verbesserte Stoffproduktion und für eine umweltverträglichere Landwirtschaft.

Das IPK Gatersleben verfügt über eine der größten Genbanken mit Ressourcen aus über 3000 botanischen Arten. Insgesamt umfasst die Sammlung des Institutes über 148000 Kulturpflanzenmuster.

Ebenfalls versucht das IPK wissenschaftliche Fragestellungen der Biologie zu Kulturpflanzen zu klären. Dies umfasst Grundlagenforschung und Anwendungen im Bereich der Züchtung resistenter und nährstoffreicher Pflanzen.

Die Aufgaben des IPK verteilen sich auf vier Abteilungen:

Abteilung Genbank

Abteilung Cytogenetik und Genomanalyse

Abteilung Molekulare Zellbiologie.

Abteilung Molekulare Genetik

Der Bereich Bioinformatik setzt sich aus Arbeitsgruppen der vier Abteilungen zusammen, um spezifische Analyse- und Modellierungsaufgaben zu lösen. Er umfasst die Arbeitsgruppen Pflanzenbioinformatik, Bioinformatik- und Informationstechnologie, Genbankdokumentation, Systembiologie und Dateninspektion.

Die Arbeitsgruppe Dateninspektion, die mit der Betreuung meiner Lernleistung vertraut ist, ist eine durch das Land Sachsen-Anhalt geförderte Nachwuchsforschergruppe mit drei Wissenschaftlern, mit Lehrerfahrungen an den Universitäten Halle und Osnabrück. Schwerpunktthema der von Herrn Dr. Marc Strickert geleiteten Gruppe ist die Verarbeitung von biologischen Hochdurchsatzdaten aus Genexpressionsexperimenten und Sequenzierstudien.

Problemstellung

Tabellengestützte Datenauswertung ist essenzieller Bestandteil der Auswertung von erhobenen Beobachtungs- und Messdaten. Am Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung in Gatersleben (IPK) nimmt die computergestützte Analyse genetischer, stoffwechselbezogener und erscheinungsbildrelevanter Merkmale einen großen Stellenwert bei der Unterstützung biologischer Hypothesenbildung und -prüfung ein. Als weit verbreitetes Werkzeug zur Datenorganisation und erstes Auswertungsmittel wird dazu die zellenwertbasierte Tabellenkalkulation Microsoft Excel verwendet.
Meine Aufgabe umfasst die Erstellung einer Microsoft Excel Anwendung zur Einfärbung von Genexpressionstabellen nach bestimmten Kriterien und die Sortierung mehrdimensionaler (vektorieller) Daten entlang von Zeilen oder Spalten innerhalb von Microsoft Excel.
Viele biologische Datenquellen liegen bereits in Tabellenform vor, aber Microsoft Excel bietet keine integrierte Funktion, die eine wertebereichsabhängige Einfärbung bzw. Sortierung dieser komplexen vektoriellen Daten übernimmt. Durch eine einfache Benutzeroberfläche werden diese Funktionen einem Anwender zur Verfügung gestellt.
Zur Erweiterung von Excel um einen Algorithmus für selbstorganisierende Merkmalskarten (SOM) wird die Standard-Programmiersprache „Visual Basic for Applications“ (VBA) genutzt.

Was ist eine Selbstorganisierende Merkmalskarte?

Eine Selbstorganisierende Merkmalskarte (SOM) ist ein Werkzeug zur Gruppierung von Datenpunkten, die durch k Attribute als Vektoren beschrieben werden [2]. Ähnliche Datenvektoren werden auf denselben oder auf benachbarte Datenrepräsentanten abgebildet, die in einem regelmäßigen Gitter angeordnet sind [3]. Im üblichen Fall eines Rechteckgitters, wie bei Kreuzungspunkten auf einem Schachbrett, kann man dann wie von einer Landkarte sprechen, auf der hochdimensionale Datenpunkte nach Ähnlichkeit gruppiert vorliegen. Ein Beispiel sind Textfragmente, die nach Vorverarbeitung thematische Inseln auf einer solchen Karte bilden und für themenbezogene Navigation in blogs und newsgroup Texten genutzt werden können (z.B. http://websom.hut.fi/websom/). In der Bioinformatik spielen SOMs bei der Gruppierung von Experimenten eine wichtige Rolle, bei denen der Aktivitätsstatus von bis zu mehreren tausend Genen untersucht werden, die dabei in Gruppen ähnlicher Genexpressionsprofile sortiert werden müssen. Damit kann dann eine Zuordnung genetischer Ähnlichkeiten zum bestimmten äußerlichen Erscheinungsformen erfolgen und Aufschluss über genetische Grundlagen von Organismen bringen.

Die Realisierung einer SOM folgt einfachen Prinzipien. Die zu untersuchenden Daten werden als Trainingsmuster behandelt, die dem Computerprogramm nach und nach in zufälliger Reihenfolge angeboten werden. Dort wird dieser Eingabereiz genutzt, um den gegenwärtigen Zustand der Datenrepräsentanten zu verbessern. Der ähnlichste Datenrepräsentant (Prototyp) wird dem angebotenen Datenpunkt noch etwas ähnlicher gemacht, also der Prototypen-Zustand in Richtung des Datenreizes verändert. Dieser Prototyp informiert ferner seine Gitternachbarn, die ihm selbst und damit auch dem angebotenen Datenvektor ähnlich sind. Diese informierten Nachbarn machen sich dem Datenvektor ebenfalls aber weniger dem Datenpunkt ähnlicher und informieren ihre weiter außen liegenden Nachbarn usw. Die entferntesten Gitterpunkte sind von der Aktualisierung in diese Richtung so gut wie nicht mehr betroffen. Wird der komplette Datensatz der Merkmalskarte mehrfach auf diese Weise präsentiert, kommt es zu einer flächigen Spezialisierung der

Datenrepräsentanten im Gitter. Der Informationsaustausch zwischen Nachbarn und die Aktualisierung der Datenrepräsentanten wird als Selbstorganisation bezeichnet.

In der vorliegenden Arbeit wird ein flächiges Gitter durch eine einfache Kette jeweils benachbarter Datenrepräsentanten ersetzt, um eine Abbildung von Datenvektoren auf diese Kette zu erzwingen, entlang deren nummerierten Datenrepräsentanten eine Sortierung erfolgen kann. In einer Tabellen-kalkulationsanwendung entspricht das der Sortierung von mehrdimensionalen Daten entlang der Zeilen, wodurch der ursprüngliche Datensatz viel übersichtlicher wird, wenn man sich durch eine so sortierte Tabelle bewegt. Man gelangt so zu einem vereinfachten Umgang mit großen Datentabellen.

Visual Basic for Applications

Visual Basic for Applications, auch VBA genannt ist eine Skriptsprache die zu den Officeprogrammen von Microsoft gehört. VBA ist von Visual Basic abgeleitet und dadurch können können bestimmte Abläufe innerhalb der Officeprogramme gesteuert werden [4][5].
VBA ist die Weiterentwicklung der bis Mitte der 90er Jahre in den Officeanwendungen enthaltenen Makrosprachen.
Jedoch ist VBA in der Funktionalität gegenüber von Visual Basic deutlich eingeschränkt.
In jeder der Officeanwendungen ist ein Visual Basic Editor vorhanden, womit es möglich ist, innerhalb der einzelnen Anwendungen komplette eigens erstellte Programme ablaufen zu lassen. Diese kann man dann z.B. auch mit Symbolleisten verknüpfen, um durch einen Knopfdruck das Programm zu starten.

Einfache VBA Grundstrukturen

Variablen

Variablen sind Platzhalter in einem Programm, denen bestimmte Werte übergeben werden können. Dabei sind die Werte einer Variablen veränderbar. Bei der Vereinbarung solcher Platzhalter sollten sinnvolle Namen verwendet werden. So wird die Lesbarkeit des Quelltextes gewährleistet. Außerdem muss für jede Variable ein Datentyp angegeben werden. Diese zuvor vereinbarten Variablen können dann in Berechnungen verwendet werden.

Datentypen

Datentyp	**Art**	**Wertebereich**
Byte	Ganze Zahlen	0...255
Integer*	Ganze Zahlen	-32.768 bis 32.767
Long	Ganze Zahlen	-2.147.483.648 bis 2.147.483.647

Single	Dezimalzahlen	Insgesamt 8 Vor- und Nachkommastellen
Double*	Dezimalzahlen	Insgesamt 16 Stellen
Currency	Kommazahlen für Währung	15 Vor- und 4 Nachkommastellen
Boolean*	Wahrheitswerte	TRUE oder FALSE
Date	Datums-Zeit-Werte	01.01.100 bis 31.12.9999
Object	Objektvariable	Verweist auf ein Objekt
String*	Text	Kann über 2Millarden Zeichen enthalten

* wird im Programm verwendet

Datenfelder

Ein Datenfeld, auch Array genannt, ist eine Zusammenfassung mehrerer Variablen des gleichen Datentyps unter einem gemeinsamen Namen. Die einzelnen Feldelemente können dann über ihren Index, in Microsoft Excel ihre Spalten- und Zeilennummer, aufgerufen werden.

Die Vereinbarung eines Arrays wird durch das Wort "Dim" eingeleitet und beinhaltet weiterhin den Namens des Arrays, in Klammern die gewünschten Dimensionen und den Datentyp der Elemente des Datenfeldes z.B.

Dim Feld (0 To 10,0 To 10) as Integer.

Damit wird ein Feld vereinbart, dass zwei Dimensionen beinhaltet, welche jeweils 10 Elemente enthält. Wir haben nun ein 10x10-Feld oder, wenn man es auf Excel bezieht, ein Feld aus 10 Zeilen und 10 Spalten.

Objekte

Microsoft Excel enthält über 150 Objekte. Als Objekt bezeichnet man alles, was programmiert und kontrolliert werden kann. Die Arbeitsmappe ist z.B. ein wichtiges Objekt in Excel. Eine Arbeitsmappe hat beispielsweise einen Autor, einen bestimmten Namen und kann auch durch ein Passwort geschützt werden. Auf solche Objekte lassen sich Methoden anwenden.

Methoden

Den Vorgang, der von bzw. mit einem Objekt ausgeführt werden kann, bezeichnet man als Methode. Eine Methode bezieht sich immer auf das davor stehende Objekt und ist von diesem Objekt durch einen Punkt getrennt.

Methode	**Beschreibung**
Select*	Markiert ein Objekt
Copy	Kopiert ein Objekt
PasteSpecial	Füllt Inhalte entsprechend der Argumente ein
ClearContents	Löscht Eingaben des angegebenen Bereiches
PrintOut	Druckt angegebene Objekte
PrintPreview	Zeigt die Seitenansicht des angegebenen Objekts

* wird im Programm verwendet

Eigenschaften

Eigenschaften beschreiben Objekte näher und geben beispielsweise den Namen, Farbe und Größe eines Objektes wieder.

Eigenschaft	**Bedeutung**
Font	Beschreibt die Schriftart des angegebenen Objekts
Size	Legt die Schriftgröße fest
Selection	Gibt das ausgewählte Objekt im ausgewählten Fenster zurück

Kontrollstrukturen

If-Then-Verzweigung

```
If    Bedingung 1 Then
      Anweisungen
ElseIf Bedingung 2 Then
      Anweisungen
Else
      Anweisung
End If
```

Für das Auswerten von einer oder wenigen Bedingungen wird die If-Then-Verzweigung verwendet. Wenn eine Bedingung erfüllt ist, wird der

dazugehörige Anweisungsblock ausgeführt. Ist keine der Bedingungen erfüllt, so wird der Anweisungsblock hinter Else (dt.:andernfalls) ausgeführt.

Select-Case-Verzweigung

```
Select Case (Variable)
      Case Variablenwert1: Anweisungen
      Case Variablenwert2: Anweisungen
      Case Variablenwert3: Anweisungen
      Case Else Anweisungen
End Select
```

Für die Auswertung von einer größeren Anzahl von alternativen Möglichkeiten verwendet man die Select-Case-Verzweigung. Während bei einer If-Then-Verzweigung eine Bedingung auf Ihren Wahrheitswert überprüft wird, erfolgt hier die Übergabe einer Variablen, deren aus dem Programmablauf resultierender Wert in mehreren Case-Zweigen mit „Variablenwert x“ verglichen wird. Bei einer Übereinstimmung wird der dazugehörige Anweisungsblock ausgeführt.
Ist jedoch keine Übereinstimmung vorhanden, so wird die Anweisung hinter „Case Else“ (dt. andernfalls) ausgeführt.

Schleife mit Vorabprüfung

```
Do While Bedingung
      Anweisungen
Loop
```

Der Anweisungsblock wird ausgeführt, solange die Bedingung erfüllt ist. Die Bedingung wird vor dem Durchlaufen der Schleife überprüft. Ist die Bedingung nicht erfüllt, so wird die Schleife übersprungen bzw. verlassen.

Schleife mit Endprüfung

```
Do
      Anweisungen
Loop Until Bedingung
```

Die Anweisungen werden solange wiederholt, bis die Bedingung erfüllt ist. Die Schleife muss mindestens einmal durchlaufen werden, weil erst nach dem Durchlaufen der Schleife überprüft wird, ob die Bedingung bereits erfüllt ist.

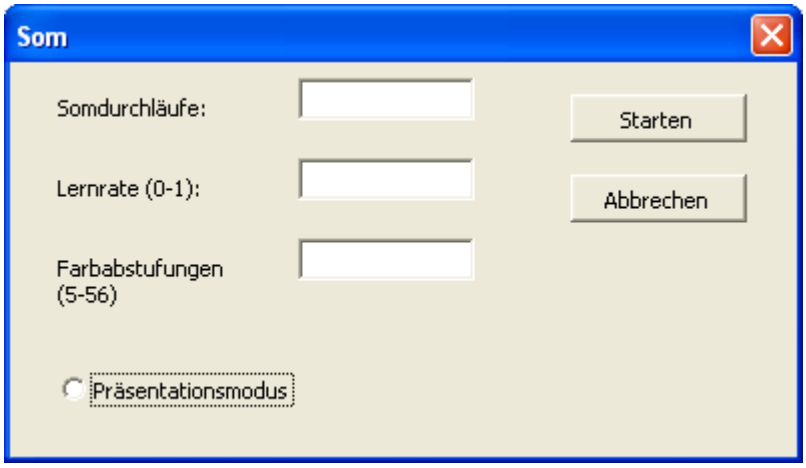
For-Next-Schleife

```
For i = AW to EW Step SW
      Anweisungen
Next i
```

Eine For-Next-Schleife wird verwendet, um eine Aktion wiederholt auszuführen. Dabei ist die Anzahl der Wiederholungen durch den Anfangswert (AW) , den Endwert (EW) und die Schrittweite (SW) festgelegt.

Aufbau des Programmes

Zu Beginn der Verwendung des Programmes muss der Benutzer in seiner Tabelle die zu sortierenden bzw. einzufärbenden Werte durch eine Selektion markieren.
Das Programm kann in der Menüzeile von Microsoft Excel unter dem Menüpunkt Lernleistung=>Som gestartet werden. Dann werden mit der abgebildeten Eingabemaske drei Programmparameter vom Benutzer abgefragt: die Anzahl der SOM-Durchläufe, die Lernrate und die Zahl der Farbabstufungen.



Weiterhin gibt es mit der Eingabemaske die Möglichkeit einen Präsentationsmodus einzuschalten. Dadurch werden zusätzlich zu den für mein Programm nötigen Ausgaben noch einige Ausgaben zur besseren Veranschaulichung ausgegeben.

Durch das Betätigen des Knopfes „Starten“ wird das Programm gestartet. Als erstes wird eine Zeilen- und Spaltenstatistik der Selektion erstellt. Anschließend werden durch das Unterprogramm „Protoinit“ erst die Zellinhalte des markierten Feldes in ein Datenfeld (Array) gespeichert, um einen schnelleren Programmzugriff auf die Daten zu gewährleisten. Weiterhin wird durch dieses Unterprogramm das Minimum und Maximum jeder Spalte des Arrays zur Initialisierung der Prototypen bestimmt. Gleichzeitig wird ein zweites Datenfeld

generiert. Dieses Feld besteht aus genauso vielen Zeilen und Spalten wie die Selektion. Es wird spaltenweise mit Zufallszahlen im Bereich zwischen dem Minimum und Maximum der entsprechenden Spalten in der Ausgangsselektion gefüllt. Dieses anfänglich mit Zufallszahlen gefüllte Feld dient dann später als Vergleichstabelle für die Originaldaten.

Anschließend wird der Euklidische Abstand der Zeilenvektoren der Originaltabelle zu den Zeilenvektoren der zufällig initialisierten Tabelle bestimmt. Dieser ergibt sich aus des Wurzel der summierten Quadrate des Zeileninhaltes nach dem Satz des Pythagoras. Diese Distanzbestimmung wird für jede Zeile des Originalfeldes zu jeder Zeile des Zufallsfeldes durchgeführt und anschließend der Gewinnerindex bestimmt, also die Zeile mit dem kleinsten Abstand zueinander. So erhält man für jede Zeile des Originalfeldes die Distanzen zu allen Zeilen des Zufallsfeldes. Wenn der Präsentationsmodus zu Beginn des Programmes aktiviert wurde, wird für jeden Zeilenvektor die Distanz, und zu einem späteren Zeitpunkt noch einmal die Distanz, der Gewinnerindex und die kleinste Distanz ausgegeben.

Anschließend werden die Prototypen noch entsprechend ihrer Nachbarschaft angepasst, d.h., dass alle Prototypen (anfänglich Zufallszahlen) entsprechend ihrer Entfernung zum Gewinnerprototypen unterschiedlich stark angenähert werden: je kleiner die Entfernung zum Gewinnerprototypen desto größer ist die Annäherung bzw. die Ortsveränderung. Je öfter diese Berechnung der Gewinnerprototypen und die Annäherung der Nachbarn wiederholt wird, desto genauer und eindeutiger wird das Ergebnis.

Der endgültig ermittelte Index des Gewinnerprototypen wird dann anschließend in einer zusätzlichen Spalte hinter der Selektion ausgegeben.

Für die Einfärbung der einzelnen Feldelemente stehen in Microsoft Excel 56 verschiedene Farben zur Verfügung.

Um eine Zuweisung einer Farbe zu einem bestimmten Zellwert mit optimaler Schrittweite zu erhalten, wird zuerst der Bereich zwischen Minimum und Maximum auf die gewünschten Anzahl der Farbabstufungen aufgeteilt. Um die jeweilige Farbe eines Feldelements zu erhalten, wird das Minimum vom Wert des jeweiligen Feldelements subtrahiert und durch die zuvor erhaltene Schrittweite geteilt nach der Formel: „Zahl = (Wert – Minimum) / Schrittweite“

Nach dem Abtrennen der Nachkommastellen erhält man eine Zahl zwischen 0

und 55, der jeweils eine bestimmte Farbe zugeordnet ist. Dadurch erhält man eine automatische kontrastreiche Einfärbung der jeweiligen Feldelemente nach folgendem Farbschema:

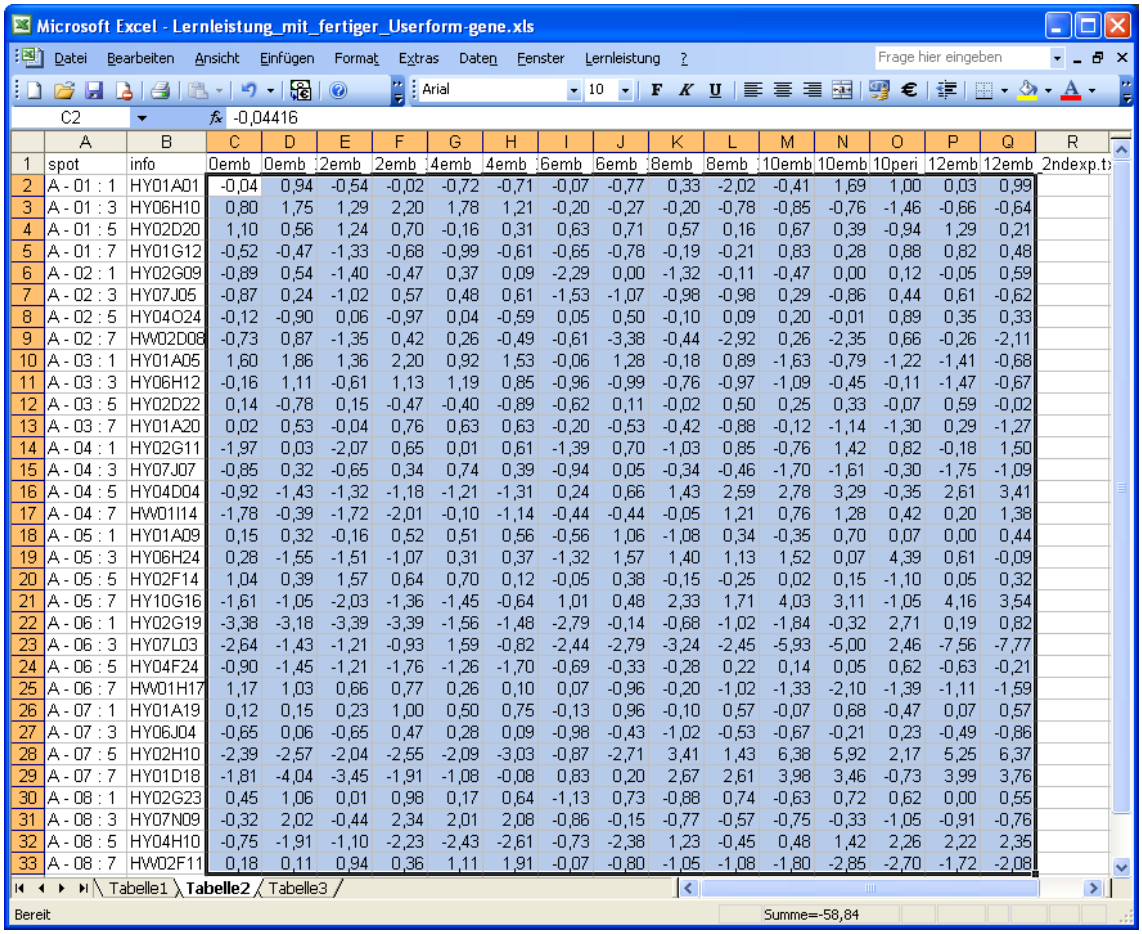
Durch das Drücken des Knopfes „Abbrechen“ wird die Eingabemaske geschlossen.

Danach empfiehlt es sich, die Tabelle entsprechend der ermittelten Gewinnerprototypenindizes zu sortieren. Dazu wird eine neue Selektion über die Tabelle und die neu hinzugefügte Spalte vom Benutzer erstellt und alle Zeilen unter dem Menüpunkt Daten > Sortieren (ohne Überschrift) nach der letzten Spalte sortiert.

Anwendungsbeispiele

Genexpressionsdaten

In dem vorliegenden Beispiel handelt es sich um eine Tabelle von Genexpressionsdaten der Gerstenkornentwicklung zu 15 verschiedenen Entwicklungszeitpunkten.



Microsoft Excel - Lernleistung_mit_fertiger_Userform-gene.xls

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster Lernleistung ?

Frage hier eingeben

Arial 10

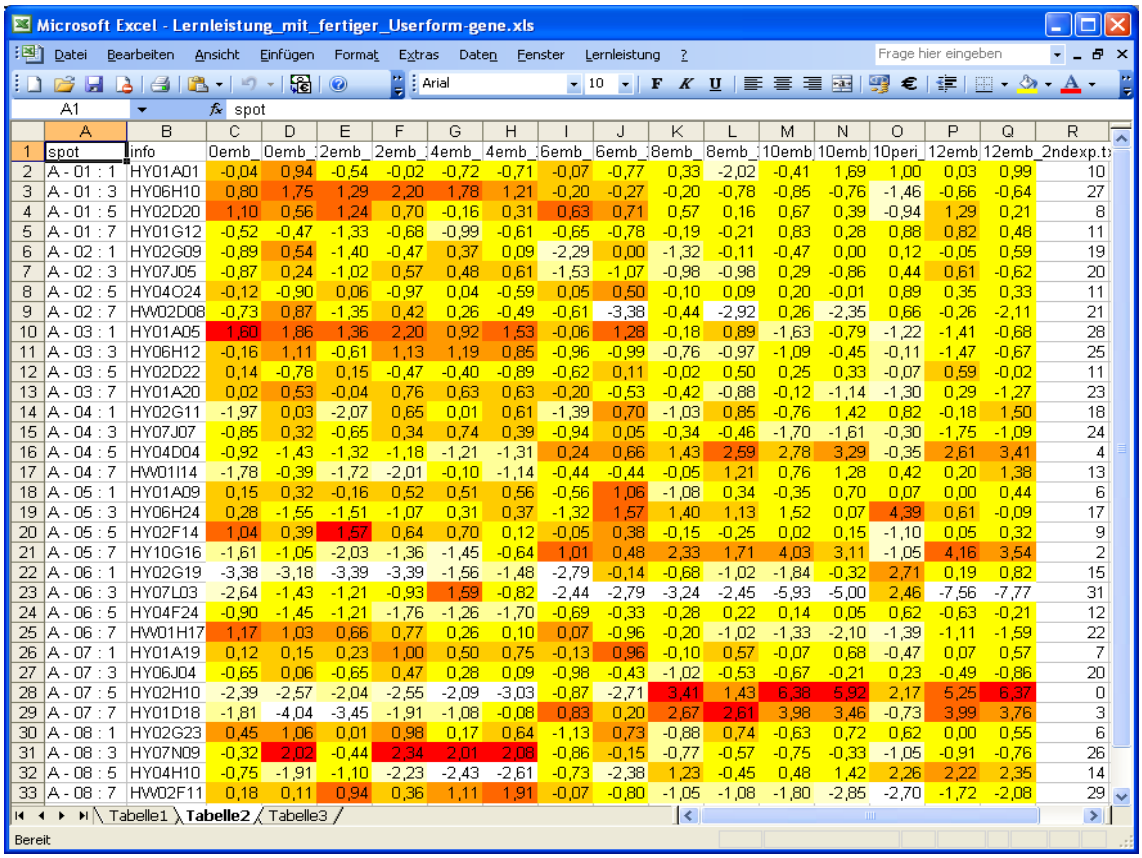
C2 -0,04416

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	spot	info	0emb	0emb	2emb	2emb	4emb	4emb	6emb	6emb	8emb	8emb	10emb	10emb	10peri	12emb	12emb	2ndexp.t
2	A - 01 : 1	HY01A01	-0,04	0,94	-0,54	-0,02	-0,72	-0,71	-0,07	-0,77	0,33	-2,02	-0,41	1,69	1,00	0,03	0,99	
3	A - 01 : 3	HY06H10	0,80	1,75	1,29	2,20	1,78	1,21	-0,20	-0,27	-0,20	-0,78	-0,85	-0,76	-1,46	-0,66	-0,64	
4	A - 01 : 5	HY02D20	1,10	0,56	1,24	0,70	-0,16	0,31	0,63	0,71	0,57	0,16	0,67	0,39	-0,94	1,29	0,21	
5	A - 01 : 7	HY01G12	-0,52	-0,47	-1,33	-0,68	-0,99	-0,61	-0,65	-0,78	-0,19	-0,21	0,83	0,28	0,88	0,82	0,48	
6	A - 02 : 1	HY02G09	-0,89	0,54	-1,40	-0,47	0,37	0,09	-2,29	0,00	-1,32	-0,11	-0,47	0,00	0,12	-0,05	0,59	
7	A - 02 : 3	HY07J05	-0,87	0,24	-1,02	0,57	0,48	0,61	-1,53	-1,07	-0,98	-0,98	0,29	-0,86	0,44	0,61	-0,62	
8	A - 02 : 5	HY04O24	-0,12	-0,90	0,06	-0,97	0,04	-0,59	0,05	0,50	-0,10	0,09	0,20	-0,01	0,89	0,35	0,33	
9	A - 02 : 7	HW02D08	-0,73	0,87	-1,35	0,42	0,26	-0,49	-0,61	-3,38	-0,44	-2,92	0,26	-2,35	0,66	-0,26	-2,11	
10	A - 03 : 1	HY01A05	1,60	1,86	1,36	2,20	0,92	1,53	-0,06	1,28	-0,18	0,89	-1,63	-0,79	-1,22	-1,41	-0,68	
11	A - 03 : 3	HY06H12	-0,16	1,11	-0,61	1,13	1,19	0,85	-0,96	-0,99	-0,76	-0,97	-1,09	-0,45	-0,11	-1,47	-0,67	
12	A - 03 : 5	HY02D22	0,14	-0,78	0,15	-0,47	-0,40	-0,89	-0,62	0,11	-0,02	0,50	0,25	0,33	-0,07	0,59	-0,02	
13	A - 03 : 7	HY01A20	0,02	0,53	-0,04	0,76	0,63	0,63	-0,20	-0,53	-0,42	-0,88	-0,12	-1,14	-1,30	0,29	-1,27	
14	A - 04 : 1	HY02G11	-1,97	0,03	-2,07	0,65	0,01	0,61	-1,39	0,70	-1,03	0,85	-0,76	1,42	0,82	-0,18	1,50	
15	A - 04 : 3	HY07J07	-0,85	0,32	-0,65	0,34	0,74	0,39	-0,94	0,05	-0,34	-0,46	-1,70	-1,61	-0,30	-1,75	-1,09	
16	A - 04 : 5	HY04D04	-0,92	-1,43	-1,32	-1,18	-1,21	-1,31	0,24	0,66	1,43	2,59	2,78	3,29	-0,35	2,61	3,41	
17	A - 04 : 7	HW01I14	-1,78	-0,39	-1,72	-2,01	-0,10	-1,14	-0,44	-0,44	-0,05	1,21	0,76	1,28	0,42	0,20	1,38	
18	A - 05 : 1	HY01A09	0,15	0,32	-0,16	0,52	0,51	0,56	-0,56	1,06	-1,08	0,34	-0,35	0,70	0,07	0,00	0,44	
19	A - 05 : 3	HY06H24	0,28	-1,55	-1,51	-1,07	0,31	0,37	-1,32	1,57	1,40	1,13	1,52	0,07	4,39	0,61	-0,09	
20	A - 05 : 5	HY02F14	1,04	0,39	1,57	0,64	0,70	0,12	-0,05	0,38	-0,15	-0,25	0,02	0,15	-1,10	0,05	0,32	
21	A - 05 : 7	HY10G16	-1,61	-1,05	-2,03	-1,36	-1,45	-0,64	1,01	0,48	2,33	1,71	4,03	3,11	-1,05	4,16	3,54	
22	A - 06 : 1	HY02G19	-3,38	-3,18	-3,39	-3,39	-1,56	-1,48	-2,79	-0,14	-0,68	-1,02	-1,84	-0,32	2,71	0,19	0,82	
23	A - 06 : 3	HY07L03	-2,64	-1,43	-1,21	-0,93	1,59	-0,82	-2,44	-2,79	-3,24	-2,45	-5,93	-5,00	2,46	-7,56	-7,77	
24	A - 06 : 5	HY04F24	-0,90	-1,45	-1,21	-1,76	-1,26	-1,70	-0,69	-0,33	-0,28	0,22	0,14	0,05	0,62	-0,63	-0,21	
25	A - 06 : 7	HW01H17	1,17	1,03	0,66	0,77	0,26	0,10	0,07	-0,96	-0,20	-1,02	-1,33	-2,10	-1,39	-1,11	-1,59	
26	A - 07 : 1	HY01A19	0,12	0,15	0,23	1,00	0,50	0,75	-0,13	0,96	-0,10	0,57	-0,07	0,68	-0,47	0,07	0,57	
27	A - 07 : 3	HY06J04	-0,65	0,06	-0,65	0,47	0,28	0,09	-0,98	-0,43	-1,02	-0,53	-0,67	-0,21	0,23	-0,49	-0,86	
28	A - 07 : 5	HY02H10	-2,39	-2,57	-2,04	-2,55	-2,09	-3,03	-0,87	-2,71	3,41	1,43	6,38	5,92	2,17	5,25	6,37	
29	A - 07 : 7	HY01D18	-1,81	-4,04	-3,45	-1,91	-1,08	-0,08	0,83	0,20	2,67	2,61	3,98	3,46	-0,73	3,99	3,76	
30	A - 08 : 1	HY02G23	0,45	1,06	0,01	0,98	0,17	0,64	-1,13	0,73	-0,88	0,74	-0,63	0,72	0,62	0,00	0,55	
31	A - 08 : 3	HY07N09	-0,32	2,02	-0,44	2,34	2,01	2,08	-0,86	-0,15	-0,77	-0,57	-0,75	-0,33	-1,05	-0,91	-0,76	
32	A - 08 : 5	HY04H10	-0,75	-1,91	-1,10	-2,23	-2,43	-2,61	-0,73	-2,38	1,23	-0,45	0,48	1,42	2,26	2,22	2,35	
33	A - 08 : 7	HW02F11	0,18	0,11	0,94	0,36	1,11	1,91	-0,07	-0,80	-1,05	-1,08	-1,80	-2,85	-2,70	-1,72	-2,08	

Tabelle1 Tabelle2 Tabelle3

Bereit Summe=-58,84

Nachdem das Programm die Exceltabelle in seinen internen Speicher (Array) übernommen hat, werden alle Distanzen zwischen den Datenzeilen und den zufällig erzeugten Prototypen berechnet und so der SOM-Algorithmus gestartet. Am Ende hat eine Spezialisierung der Gewinnerprototypen auf einzelne Datenzeilen stattgefunden und deren Index wird in einer weiteren Spalte ausgegeben. Anschließend erfolgt die Einfärbung der Datenfelder in Abhängigkeit des Wertebereiches der einzelnen Spalten wie im Folgenden illustriert.



Microsoft Excel - Lernleistung_mit_fertiger_Userform-gene.xls

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster Lernleistung ?

Frage hier eingeben

Arial 10

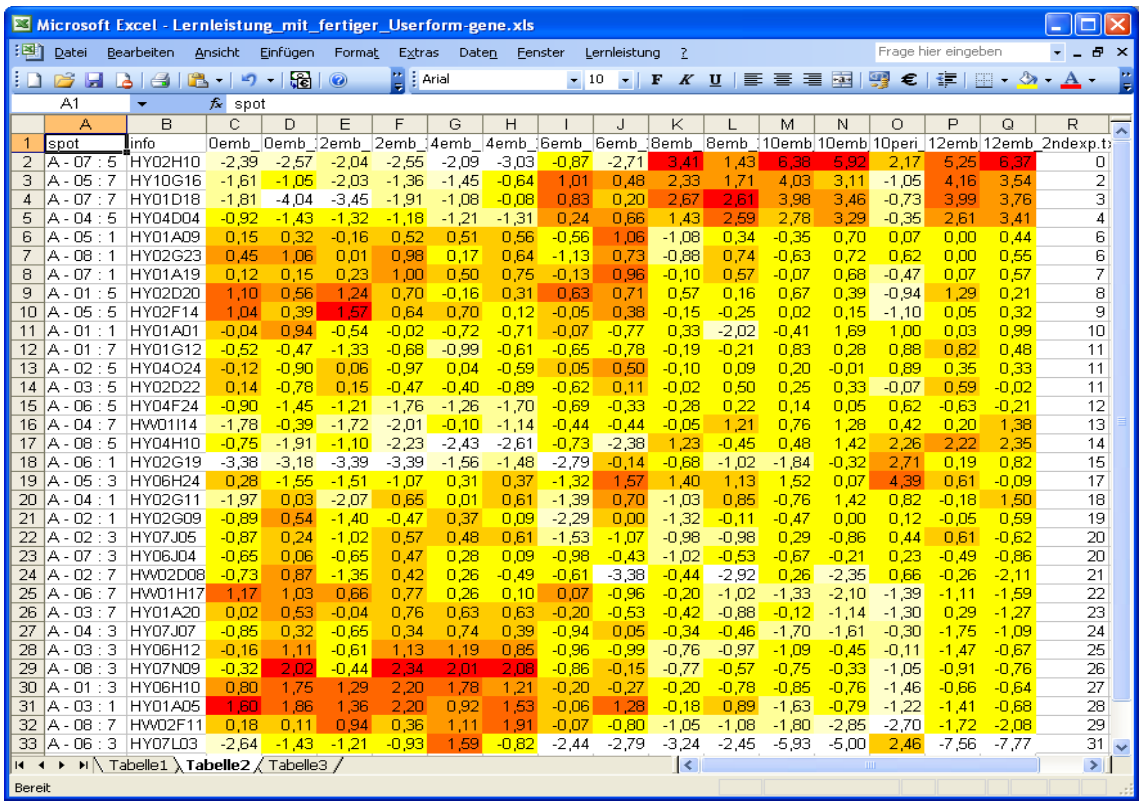
A1 spot

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	spot	info	0emb_	0emb_	2emb_	2emb_	4emb_	4emb_	6emb_	6emb_	8emb_	8emb_	10emb	10emb	10peri_	12emb	12emb_	2ndexp.t
2	A - 01 : 1	HY01A01	-0,04	0,94	-0,54	-0,02	-0,72	-0,71	-0,07	-0,77	0,33	-2,02	-0,41	1,69	1,00	0,03	0,99	10
3	A - 01 : 3	HY06H10	0,80	1,75	1,29	2,20	1,78	1,21	-0,20	-0,27	-0,20	-0,78	-0,85	-0,76	-1,46	-0,66	-0,64	27
4	A - 01 : 5	HY02D20	1,10	0,56	1,24	0,70	-0,16	0,31	0,63	0,71	0,57	0,16	0,67	0,39	-0,94	1,29	0,21	8
5	A - 01 : 7	HY01G12	-0,52	-0,47	-1,33	-0,68	-0,99	-0,61	-0,65	-0,78	-0,19	-0,21	0,83	0,28	0,88	0,82	0,48	11
6	A - 02 : 1	HY02G09	-0,89	0,54	-1,40	-0,47	0,37	0,09	-2,29	0,00	-1,32	-0,11	-0,47	0,00	0,12	-0,05	0,59	19
7	A - 02 : 3	HY07J05	-0,87	0,24	-1,02	0,57	0,48	0,61	-1,53	-1,07	-0,98	-0,98	0,29	-0,86	0,44	0,61	-0,62	20
8	A - 02 : 5	HY04O24	-0,12	-0,90	0,06	-0,97	0,04	-0,59	0,05	0,50	-0,10	0,09	0,20	-0,01	0,89	0,35	0,33	11
9	A - 02 : 7	HW02D08	-0,73	0,87	-1,35	0,42	0,26	-0,49	-0,61	-3,38	-0,44	-2,92	0,26	-2,35	0,66	-0,26	-2,11	21
10	A - 03 : 1	HY01A05	1,60	1,86	1,36	2,20	0,92	1,53	-0,06	1,28	-0,18	0,89	-1,63	-0,79	-1,22	-1,41	-0,68	28
11	A - 03 : 3	HY06H12	-0,16	1,11	-0,61	1,13	1,19	0,85	-0,96	-0,99	-0,76	-0,97	-1,09	-0,45	-0,11	-1,47	-0,67	25
12	A - 03 : 5	HY02D22	0,14	-0,78	0,15	-0,47	-0,40	-0,89	-0,62	0,11	-0,02	0,50	0,25	0,33	-0,07	0,59	-0,02	11
13	A - 03 : 7	HY01A20	0,02	0,53	-0,04	0,76	0,63	0,63	-0,20	-0,53	-0,42	-0,88	-0,12	-1,14	-1,30	0,29	-1,27	23
14	A - 04 : 1	HY02G11	-1,97	0,03	-2,07	0,65	0,01	0,61	-1,39	0,70	-1,03	0,85	-0,76	1,42	0,82	-0,18	1,50	18
15	A - 04 : 3	HY07J07	-0,85	0,32	-0,65	0,34	0,74	0,39	-0,94	0,05	-0,34	-0,46	-1,70	-1,61	-0,30	-1,75	-1,09	24
16	A - 04 : 5	HY04D04	-0,92	-1,43	-1,32	-1,18	-1,21	-1,31	0,24	0,66	1,43	2,59	2,78	3,29	-0,35	2,61	3,41	4
17	A - 04 : 7	HW01I14	-1,78	-0,39	-1,72	-2,01	-0,10	-1,14	-0,44	-0,44	-0,05	1,21	0,76	1,28	0,42	0,20	1,38	13
18	A - 05 : 1	HY01A09	0,15	0,32	-0,16	0,52	0,51	0,56	-0,56	1,06	-1,08	0,34	-0,35	0,70	0,07	0,00	0,44	6
19	A - 05 : 3	HY06H24	0,28	-1,55	-1,51	-1,07	0,31	0,37	-1,32	1,57	1,40	1,13	1,52	0,07	4,39	0,61	-0,09	17
20	A - 05 : 5	HY02F14	1,04	0,39	1,57	0,64	0,70	0,12	-0,05	0,38	-0,15	-0,25	0,02	0,15	-1,10	0,05	0,32	9
21	A - 05 : 7	HY10G16	-1,61	-1,05	-2,03	-1,36	-1,45	-0,64	1,01	0,48	2,33	1,71	4,03	3,11	-1,05	4,16	3,54	2
22	A - 06 : 1	HY02G19	-3,38	-3,18	-3,39	-3,39	-1,56	-1,48	-2,79	-0,14	-0,68	-1,02	-1,84	-0,32	2,71	0,19	0,82	15
23	A - 06 : 3	HY07L03	-2,64	-1,43	-1,21	-0,93	1,59	-0,82	-2,44	-2,79	-3,24	-2,45	-5,93	-5,00	2,46	-7,56	-7,77	31
24	A - 06 : 5	HY04F24	-0,90	-1,45	-1,21	-1,76	-1,26	-1,70	-0,69	-0,33	-0,28	0,22	0,14	0,05	0,62	-0,63	-0,21	12
25	A - 06 : 7	HW01H17	1,17	1,03	0,66	0,77	0,26	0,10	0,07	-0,96	-0,20	-1,02	-1,33	-2,10	-1,39	-1,11	-1,59	22
26	A - 07 : 1	HY01A19	0,12	0,15	0,23	1,00	0,50	0,75	-0,13	0,96	-0,10	0,57	-0,07	0,68	-0,47	0,07	0,57	7
27	A - 07 : 3	HY06J04	-0,65	0,06	-0,65	0,47	0,28	0,09	-0,98	-0,43	-1,02	-0,53	-0,67	-0,21	0,23	-0,49	-0,86	20
28	A - 07 : 5	HY02H10	-2,39	-2,57	-2,04	-2,55	-2,09	-3,03	-0,87	-2,71	3,41	1,43	6,38	5,92	2,17	5,25	6,37	0
29	A - 07 : 7	HY01D18	-1,81	-4,04	-3,45	-1,91	-1,08	-0,08	0,83	0,20	2,67	2,61	3,98	3,46	-0,73	3,99	3,76	3
30	A - 08 : 1	HY02G23	0,45	1,06	0,01	0,98	0,17	0,64	-1,13	0,73	-0,88	0,74	-0,63	0,72	0,62	0,00	0,55	6
31	A - 08 : 3	HY07N09	-0,32	2,02	-0,44	2,34	2,01	2,08	-0,86	-0,15	-0,77	-0,57	-0,75	-0,33	-1,05	-0,91	-0,76	26
32	A - 08 : 5	HY04H10	-0,75	-1,91	-1,10	-2,23	-2,43	-2,61	-0,73	-2,38	1,23	-0,45	0,48	1,42	2,26	2,22	2,35	14
33	A - 08 : 7	HW02F11	0,18	0,11	0,94	0,36	1,11	1,91	-0,07	-0,80	-1,05	-1,08	-1,80	-2,85	-2,70	-1,72	-2,08	29

Tabelle1 Tabelle2 Tabelle3

Bereit

Anschließend kann man die Tabelle nach der letzten neu eingefügten Tabellenspalte der Gewinnerprototypen sortieren. Dadurch wird die Ähnlichkeit der koexprimierten Gene deutlich hervorgehoben.



Microsoft Excel - Lernleistung_mit_fertiger_Userform-gene.xls

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster Lernleistung ?

Frage hier eingeben

Arial 10

A1 spot

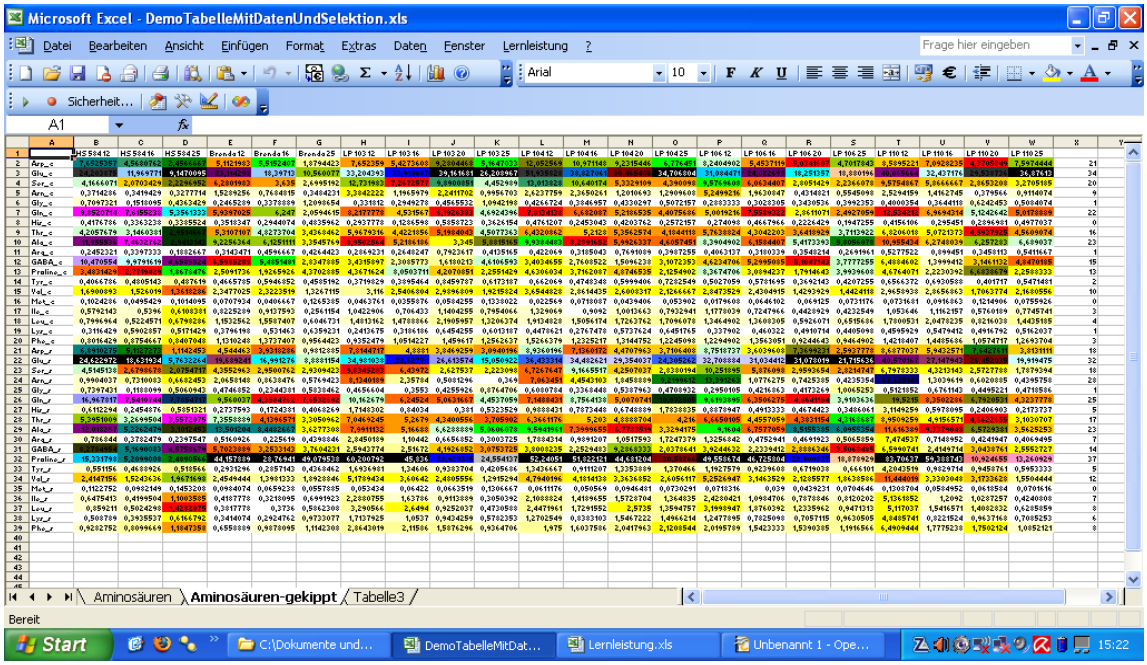
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	spot	info	0emb_	0emb_	2emb_	2emb_	4emb_	4emb_	6emb_	6emb_	8emb_	8emb_	10emb	10emb	10peri_	12emb	12emb_	2ndexp.t
2	A - 07 : 5	HY02H10	-2,39	-2,57	-2,04	-2,55	-2,09	-3,03	-0,87	-2,71	3,41	1,43	6,38	5,92	2,17	5,25	6,37	0
3	A - 05 : 7	HY10G16	-1,61	-1,05	-2,03	-1,36	-1,45	-0,64	1,01	0,48	2,33	1,71	4,03	3,11	-1,05	4,16	3,54	2
4	A - 07 : 7	HY01D18	-1,81	-4,04	-3,45	-1,91	-1,08	-0,08	0,83	0,20	2,67	2,61	3,98	3,46	-0,73	3,99	3,76	3
5	A - 04 : 5	HY04D04	-0,92	-1,43	-1,32	-1,18	-1,21	-1,31	0,24	0,66	1,43	2,59	2,78	3,29	-0,35	2,61	3,41	4
6	A - 05 : 1	HY01A09	0,15	0,32	-0,16	0,52	0,51	0,56	-0,56	1,06	-1,08	0,34	-0,35	0,70	0,07	0,00	0,44	6
7	A - 08 : 1	HY02G23	0,45	1,06	0,01	0,98	0,17	0,64	-1,13	0,73	-0,88	0,74	-0,63	0,72	0,62	0,00	0,55	6
8	A - 07 : 1	HY01A19	0,12	0,15	0,23	1,00	0,50	0,75	-0,13	0,96	-0,10	0,57	-0,07	0,68	-0,47	0,07	0,57	7
9	A - 01 : 5	HY02D20	1,10	0,56	1,24	0,70	-0,16	0,31	0,63	0,71	0,57	0,16	0,67	0,39	-0,94	1,29	0,21	8
10	A - 05 : 5	HY02F14	1,04	0,39	1,57	0,64	0,70	0,12	-0,05	0,38	-0,15	-0,25	0,02	0,15	-1,10	0,05	0,32	9
11	A - 01 : 1	HY01A01	-0,04	0,94	-0,54	-0,02	-0,72	-0,71	-0,07	-0,77	0,33	-2,02	-0,41	1,69	1,00	0,03	0,99	10
12	A - 01 : 7	HY01G12	-0,52	-0,47	-1,33	-0,68	-0,99	-0,61	-0,65	-0,78	-0,19	-0,21	0,83	0,28	0,88	0,82	0,48	11
13	A - 02 : 5	HY04O24	-0,12	-0,90	0,06	-0,97	0,04	-0,59	0,05	0,50	-0,10	0,09	0,20	-0,01	0,89	0,35	0,33	11
14	A - 03 : 5	HY02D22	0,14	-0,78	0,15	-0,47	-0,40	-0,89	-0,62	0,11	-0,02	0,50	0,25	0,33	-0,07	0,59	-0,02	11
15	A - 06 : 5	HY04F24	-0,90	-1,45	-1,21	-1,76	-1,26	-1,70	-0,69	-0,33	-0,28	0,22	0,14	0,05	0,62	-0,63	-0,21	12
16	A - 04 : 7	HW01I14	-1,78	-0,39	-1,72	-2,01	-0,10	-1,14	-0,44	-0,44	-0,05	1,21	0,76	1,28	0,42	0,20	1,38	13
17	A - 08 : 5	HY04H10	-0,75	-1,91	-1,10	-2,23	-2,43	-2,61	-0,73	-2,38	1,23	-0,45	0,48	1,42	2,26	2,22	2,35	14
18	A - 06 : 1	HY02G19	-3,38	-3,18	-3,39	-3,39	-1,56	-1,48	-2,79	-0,14	-0,68	-1,02	-1,84	-0,32	2,71	0,19	0,82	15
19	A - 05 : 3	HY06H24	0,28	-1,55	-1,51	-1,07	0,31	0,37	-1,32	1,57	1,40	1,13	1,52	0,07	4,39	0,61	-0,09	17
20	A - 04 : 1	HY02G11	-1,97	0,03	-2,07	0,65	0,01	0,61	-1,39	0,70	-1,03	0,85	-0,76	1,42	0,82	-0,18	1,50	18
21	A - 02 : 1	HY02G09	-0,89	0,54	-1,40	-0,47	0,37	0,09	-2,29	0,00	-1,32	-0,11	-0,47	0,00	0,12	-0,05	0,59	19
22	A - 02 : 3	HY07J05	-0,87	0,24	-1,02	0,57	0,48	0,61	-1,53	-1,07	-0,98	-0,98	0,29	-0,86	0,44	0,61	-0,62	20
23	A - 07 : 3	HY06J04	-0,65	0,06	-0,65	0,47	0,28	0,09	-0,98	-0,43	-1,02	-0,53	-0,67	-0,21	0,23	-0,49	-0,86	20
24	A - 02 : 7	HW02D08	-0,73	0,87	-1,35	0,42	0,26	-0,49	-0,61	-3,38	-0,44	-2,92	0,26	-2,35	0,66	-0,26	-2,11	21
25	A - 06 : 7	HW01H17	1,17	1,03	0,66	0,77	0,26	0,10	0,07	-0,96	-0,20	-1,02	-1,33	-2,10	-1,39	-1,11	-1,59	22
26	A - 03 : 7	HY01A20	0,02	0,53	-0,04	0,76	0,63	0,63	-0,20	-0,53	-0,42	-0,88	-0,12	-1,14	-1,30	0,29	-1,27	23
27	A - 04 : 3	HY07J07	-0,85	0,32	-0,65	0,34	0,74	0,39	-0,94	0,05	-0,34	-0,46	-1,70	-1,61	-0,30	-1,75	-1,09	24
28	A - 03 : 3	HY06H12	-0,16	1,11	-0,61	1,13	1,19	0,85	-0,96	-0,99	-0,76	-0,97	-1,09	-0,45	-0,11	-1,47	-0,67	25
29	A - 08 : 3	HY07N09	-0,32	2,02	-0,44	2,34	2,01	2,08	-0,86	-0,15	-0,77	-0,57	-0,75	-0,33	-1,05	-0,91	-0,76	26
30	A - 01 : 3	HY06H10	0,80	1,75	1,29	2,20	1,78	1,21	-0,20	-0,27	-0,20	-0,78	-0,85	-0,76	-1,46	-0,66	-0,64	27
31	A - 03 : 1	HY01A05	1,60	1,86	1,36	2,20	0,92	1,53	-0,06	1,28	-0,18	0,89	-1,63	-0,79	-1,22	-1,41	-0,68	28
32	A - 08 : 7	HW02F11	0,18	0,11	0,94	0,36	1,11	1,91	-0,07	-0,80	-1,05	-1,08	-1,80	-2,85	-2,70	-1,72	-2,08	29
33	A - 06 : 3	HY07L03	-2,64	-1,43	-1,21	-0,93	1,59	-0,82	-2,44	-2,79	-3,24	-2,45	-5,93	-5,00	2,46	-7,56	-7,77	31

Tabelle1 Tabelle2 Tabelle3

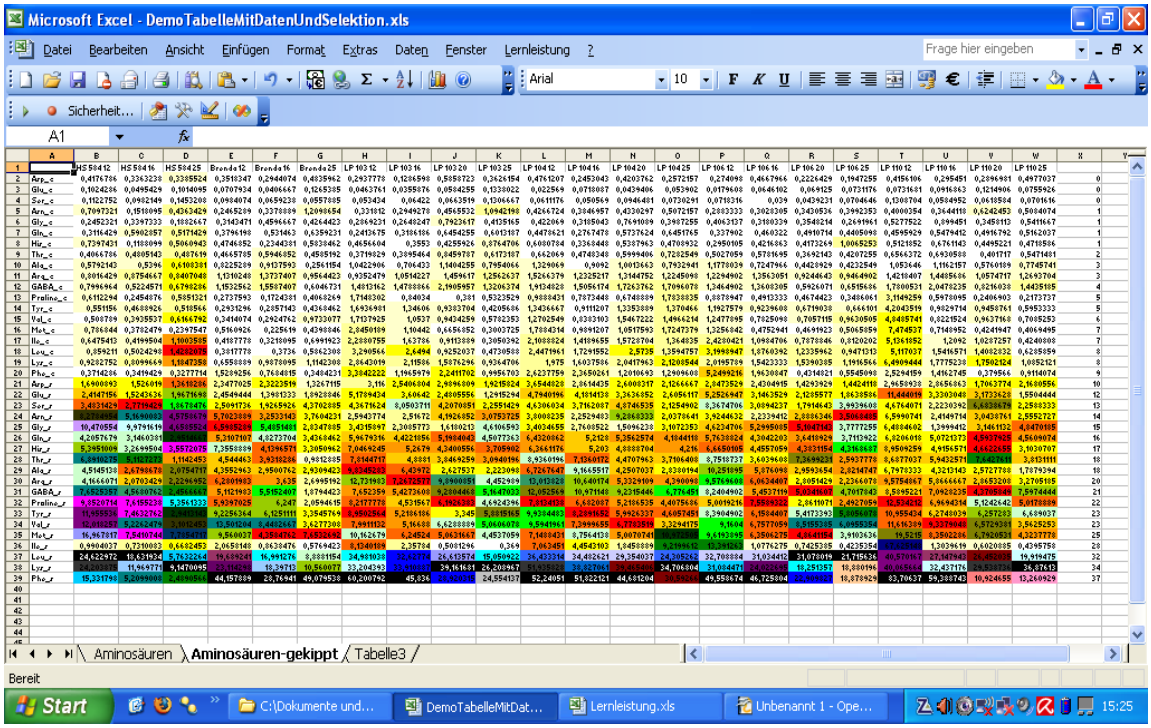
Bereit

Aminosäuremessungen

Auch im Falle von Messungen von 22 verschiedenen Aminosäuren in verschiedenen Gerstensorten kann das Programm erfolgreich angewendet werden. Aus der folgenden unsortierten Tabelle



ergibt sich nach der Sortierung per SOM folgendes Bild, das Biologen eine leichtere Interpretation der Aminosäure-Konzentrationen ermöglicht.



Resümee

Das im Rahmen der besonderen Lernleistung erstellte Programm stellt sich für den Umgang mit mehrdimensionalen Daten, wie sie in verschiedenen Bereichen biowissenschaftlicher Datenanalyse auftreten, als hilfreich dar. Seine Hauptmerkmale sind die benutzerfreundliche Bedienung, die Sortiermöglichkeiten und die Einfärbung der Tabellenzellen entsprechend ihrer Werte.

Nach ausgiebigen Tests des Programmes wurden auch einige Schwächen erkennbar. So kommt es bei der Anwendung des Programmes auf relativ große Datensätze mit mehr als 1000 Zeilen zu langen Wartezeiten.

Weiterhin ist es aufgrund der derzeit realisierten Variableninitialisierung nicht möglich, das Programm mehrmals hintereinander mit verschiedenen Werten zu starten. Für den weiteren Benutzerkomfort wäre eine automatische Sortierung der Tabelle wünschenswert, die aber einen anderen Umgang mit dem Excel-internen Selektionsmechanismus für die Zellen erfordern würde.

Da Microsoft Excel nur 56 verschiedene Farben bereitstellt kann eine größere Farbpalette nicht angeboten werden, was aber in der Praxis keine Einschränkung der Nutzbarkeit bedeutet.

Des Weiteren darf die Tabelle nicht mehr als 26 Spalten enthalten, da bei der 27. Spalte die Bezeichnung der Spalten mit Doppelbuchstaben erfolgt, für die die aktuelle Konvertierungsroutine nicht ausgelegt ist. In meinem Programm ist bei der Umrechnung der Spaltennummer in einen Spaltenbuchstaben nur ein Buchstabe möglich.

Bei der Programmentwicklung hat sich gezeigt, dass im Falle von Programmierfehlern die Hinweise von VBA nicht immer zu einer leichten Fehlerdiagnose ausreichen.

Die Projektarbeit hat mir sowohl einen Einblick in anwendungsbezogene Programmentwicklung als auch in einige Fragestellungen biologischer Datenanalyse ermöglicht.

Anhang

Quellcode

```
Option Explicit

Public somdurchlaeufe As Integer
Public lern_rate As Double
Public presentationsmodus As Boolean
Public anzahl As Variant
Public min As Double
Public max As Double
Public n_prot As Double
Public n_row As Double
Public n_col As Double
Public Farbtabelle As Variant
Public adbProtos() As Double
Public adbFeld() As Variant
Public act_col As Integer
Public act_row As Integer
Public str_col As String
Public first_col, first_row As Double

Sub UserForm()

    Farbtabelle = Array(2, 19, 36, 6, 27, 44, 45, 46, 3, 35, 4, 43,
12, 50, 10, 14, 31, 51, 52, 38, 22, 7, 26, 39, 18, 54, 13, 29, 21, 34,
20, 28, 8, 42, 33, 37, 24, 17, 41, 32, 5, 23, 47, 55, 11, 25, 49, 40,
9, 30, 53, 15, 48, 16, 56, 1)

    UserForm1.Show

End Sub

Public Function GetExcelCol(ByVal lidx As Long, Optional ByVal
binitialCall As Boolean = True) As String

    If (binitialCall) Then lidx = lidx + 1
    If (lidx = 0) Then Exit Function
```

```
    GetExcelCol = GetExcelCol((lidx - 1) \ 26, False) + Chr(65 + (lidx
- 2) Mod 26)

End Function


Sub zaehlen()

    n_row = Selection.Rows.Count                              'Zeilen
und Spalten der Selektion zählen
    n_col = Selection.Columns.Count
    first_row = Selection.Row                                 'Start-
Zeile (links oben) der Selektion
    first_col = Selection.Column                              'Start-
Spalte (links oben) der Selektion

    MsgBox "Anzahl Zeilen: " & n_row & " / Anzahl Spalten: " & n_col
'    MsgBox "erste Zeile: " & first_row & " / erste Spalte: " &
first_col

End Sub

Sub MinimumMaximum(y As Integer)

    Dim x As Integer
    Dim zelle As Variant


    act_col = first_col + y                                   'aktive
Spalte als Zahl bestimmen

    str_col = GetExcelCol(act_col)                            'Aktive
Spalte in Buchstaben umrechnen
    act_row = first_row

    For x = 0 To n_row - 1
        adbFeld(y, x) = Range(str_col & (x + act_row))        'jede
Zelle der Spalte wird in 2-dim. Array eingelesen (Erzeugung der
Spalten-Adresse)
    Next x
```

```
    min = 10 ^ 9
    max = -10 ^ 9
    For x = 0 To n_row - 1
        zelle = adbFeld(y, x)
'Bestimmung Minimalwert
        If zelle < min Then
            min = zelle
        End If
        If zelle > max Then
            max = zelle
        End If
    Next x
    MsgBox "Das Minimum/Maximum für Spalte " & str_col & " ist: " &
min & "/" & max 'Ausgabe

End Sub

Function Distanz(protonum As Integer, datanum As Integer)

    Dim y As Integer
    Dim sm As Double

    sm = 0

    For y = 0 To n_col - 1
        sm = sm + (adbProtos(y, protonum) - adbFeld(y, datanum)) ^ 2
    Next y

    Distanz = Sqr(sm)

End Function

Sub Protoinit()

Dim y As Integer
Dim x As Integer

    For y = 1 To n_col
        MinimumMaximum (y - 1)
        For x = 1 To n_row
            adbProtos(y, x) = min + Rnd * (max - min)
```

```
        Next x
    Next y

End Sub

  Sub Main()

    zaehlen

    n_prot = n_row ' kann per GUI gesetzt werden

    Dim x As Integer
    Dim y As Integer
    Dim wert As Double
    Dim sw As Double
'Schrittweite
    Dim zahl As Double
'entstandene Zahl von der die Farbe abhängt
    Dim i As Integer
    Dim j As Integer
    Dim k As Integer
    Dim l As Integer
    Dim t As Integer
    Dim gewinnerindex As Integer
    Dim min_dist As Double
    Dim akt_dist As Double
    Dim seit_schritt As Integer
    Dim nachbar_reichw As Double
    Dim nachbar_staerke As Double
    Dim nachbar_reichweit As Double
'Startgröße der Nachbarschaft
    Dim nachbar_nah As Double                                    'Endgröße
der Nachbarschaft



    nachbar_reichw = n_prot

    ReDim Preserve adbFeld(n_col - 1, n_row - 1)
'Festlegung der Feldgröße
    ReDim Preserve adbProtos(n_col, n_prot)
```

```
    Protoinit

    For t = somdurchlaeufe To 1 Step -1

        nachbar_reichweit = nachbar_reichw * t / somdurchlaeufe

        For i = 0 To n_row - 1 ' alle Daten

            min_dist = 10 ^ 10 ' fast unendlich

            For j = 0 To n_prot - 1 ' alle Prototypen
                akt_dist = Distanz(j, i)
                If akt_dist < min_dist Then
                    min_dist = akt_dist
                    gewinnerindex = j
                End If
                If presentationsmodus Then
                    MsgBox "Distanz: " & akt_dist
                End If
            Next j
            If presentationsmodus Then
                MsgBox "Gewinner zum Datenpunkt " & i & " ist Prototyp
" & gewinnerindex & ". Distanz ist: " & min_dist
            End If

  ' passe prototypen entsprechend Gauss-gewichteter nachbarschaft an
            seit_schritt = 0
            For k = gewinnerindex To n_prot - 1 ' adaptiere alle
protos oberhalb gewinnerindex
                nachbar_staerke = lern_rate * Exp(-seit_schritt ^ 2 /
nachbar_reichweit ^ 2)
                For y = 0 To n_col - 1
                    adbProtos(y, k) = adbProtos(y, k) -
nachbar_staerke * (adbProtos(y, k) - adbFeld(y, i))
                Next y
                seit_schritt = seit_schritt + 1
            Next k

            seit_schritt = 1
            For l = gewinnerindex - 1 To 0 Step -1 ' adaptiere alle
```

```
protos echt unterhalb gewinnerindex
                nachbar_staerke = lern_rate * Exp(-seit_schritt ^ 2 /
nachbar_reichweit ^ 2)
                For y = 0 To n_col - 1
                    adbProtos(y, l) = adbProtos(y, l) -
nachbar_staerke * (adbProtos(y, l) - adbFeld(y, i))
                Next y
                seit_schritt = seit_schritt + 1
            Next l
        Next i

    Next t

 'ordne am ende daten trainierte protoypen zu

      str_col = GetExcelCol(first_col + n_col)

      For i = 0 To n_row - 1 ' alle Daten

        min_dist = 10 ^ 10 ' fast unendlich

        For j = 0 To n_prot - 1 ' alle Prototypen
          akt_dist = Distanz(j, i)
          If akt_dist < min_dist Then
            min_dist = akt_dist
            gewinnerindex = j
          End If
        Next j
        Range(str_col & i + first_row).Value = gewinnerindex
      Next i

    For y = 1 To n_col

        MinimumMaximum (y - 1)                                  'Prozedur
MinimumMaximum wird aufgerufen

        sw = (max - min) / (anzahl - 1)                          'bei
10 verschiedenen Farbwerten

        For x = 1 To n_row
```

```
            Range(Cells(first_row + x - 1, first_col + y - 1),
Cells(first_row + x - 1, first_col + y - 1)).Select      'Markiert
einen Bereich
            wert = adbFeld(y - 1, x - 1)
            zahl = (wert - min) / sw
            zahl = Fix(zahl)
            Selection.Interior.ColorIndex = Farbtabelle(zahl)
        Next x
    Next y

End Sub
```

Literaturverzeichnis

1: Webseite des IPK-Gatersleben: http://www.ipk-gatersleben.de

2: Teuvo Kohonen, Self-Organisation and Associative Memory, 1983

3: Teuvo Kohonen, Self-Organizing Maps , 2001

4: Natasch Nicol, Excel 2002/2003 programmieren, 2005

5: Michael Kofler, Excel-VBA programmieren, 2006