**Übungsblatt 9**

**Genexpressionsanalyse III: Klassifikation**

**Aufgabe 9.1**

Ein Bild, das Tisch enthält.

Automatisch generierte BeschreibungEin Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Hierbei fällt auf, dass das Markergen z für die Zurodnung des Samples zu einem gesunden oder kranken Gewebe nicht entscheidend ist, d.h. nur die Markergene x und y definieren den „tumor signature“, was daran liegt, dass es z in allen Samples (unabhängig davon, ob vom gesunden oder kranken Gewebe) den gleichen Wert zeigt. Aus diesem Grund werde ich nur sie für die Klassifikation der Samples stammend aus einem unbekannten Gewebe verwenden.

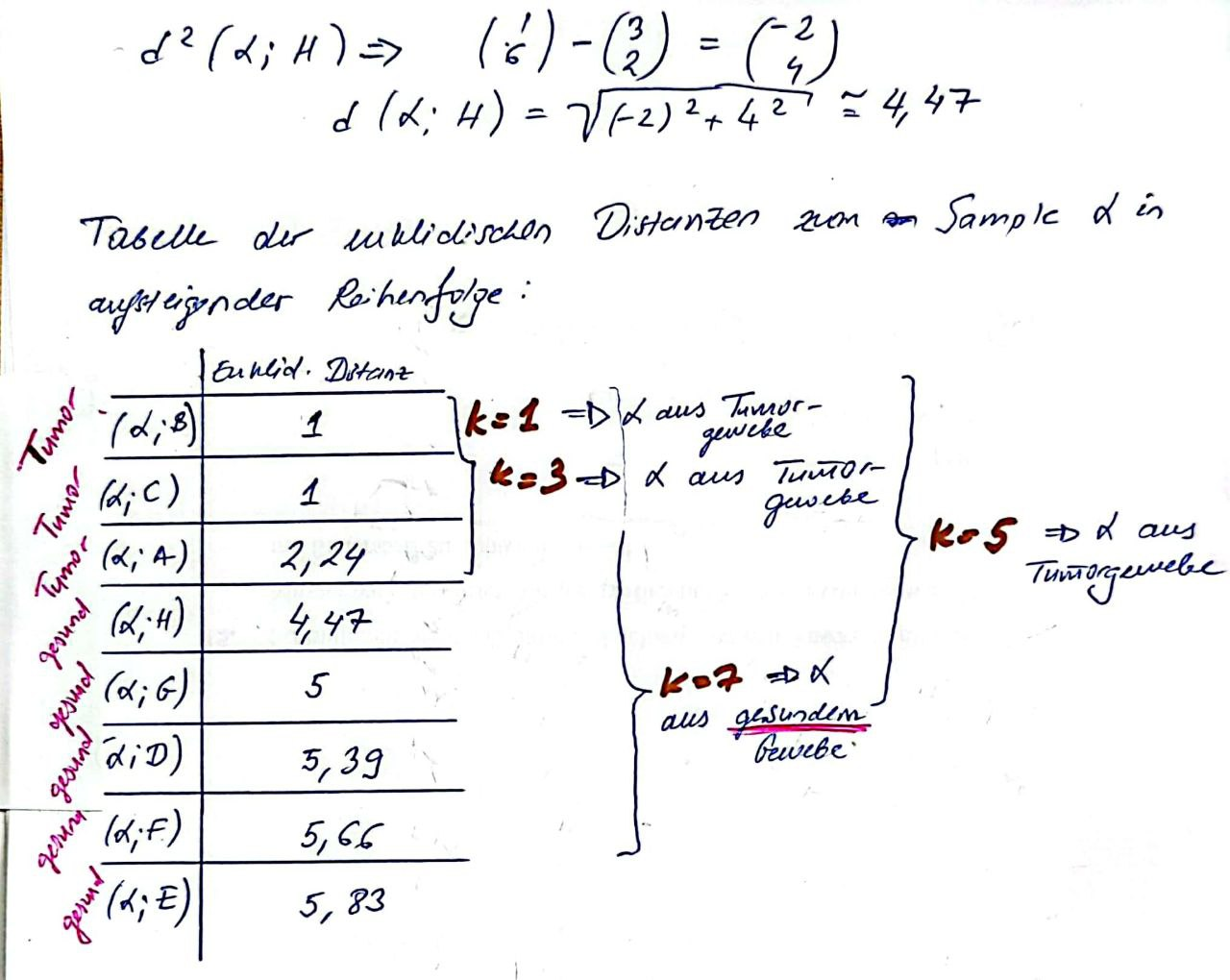
Werte der Gene aus dem unbekannten Gewebe:

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text enthält.

Automatisch generierte BeschreibungSample α:



Sample β:

Ein Bild, das Text enthält.

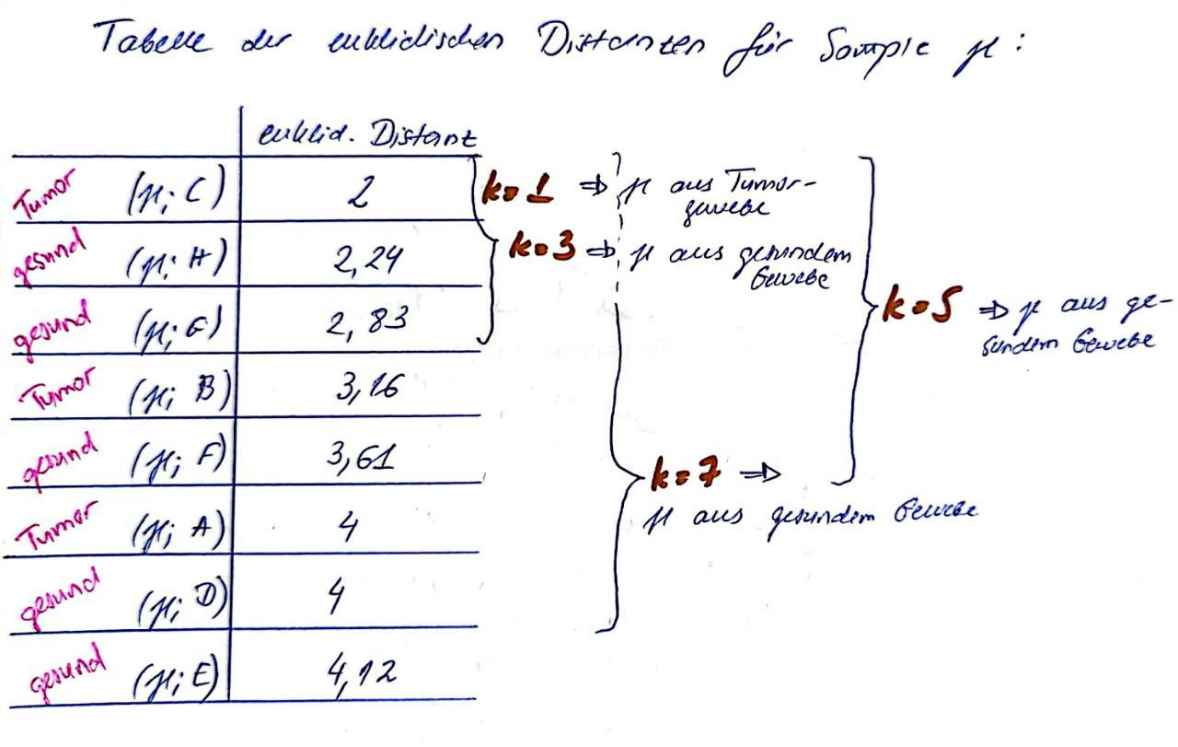
Automatisch generierte Beschreibung

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text enthält.

Automatisch generierte BeschreibungSample γ:



Interpretation der Resultate: KNN funktioniert gut, wenn man ein gutes Clustering hat. D.h. wenn die Cluster aus Datenpunkten besteht, die nah zueinander sind. Außerdem ist es von großer Bedeutung, wie groß das *k* gewählt wird, da z.B. bei kleinem *k* das Risiko, dass man ein Sample falsch zuordnet, steigt. Dies könnte daran liegen, dass der am nächsten stehende Datenpunkt ein Ausreißer ist.

**Aufgabe 9.2**

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

**Aufgabe 9.3**

> install.packages("class")

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/

Installing package into ‘C:/Users/raliz/AppData/Local/R/win-library/4.2’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/class\_7.3-20.zip'

Content type 'application/zip' length 97135 bytes (94 KB)

downloaded 94 KB

package ‘class’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\raliz\AppData\Local\Temp\RtmpC2r4lT\downloaded\_packages

> genedata <- read.csv("C:\\Users\\raliz\\Documents\\Universität\\12. Semester\\Bioinformatik\\Übungen\\Übung 9\\genedata-for-knn.csv")

> library("class")

> knn.cv(genedata, classification, 7, p = 1, classification)

Error in knn.cv(genedata, classification, 7, p = 1, classification) :

object 'classification' not found

> knn.cv(genedata, classification, 7, p = 1, "classification")

Error in knn.cv(genedata, classification, 7, p = 1, "classification") :

object 'classification' not found

> knn.cv(genedata, "classification", 7, p = 1, "classification")

Error in knn.cv(genedata, "classification", 7, p = 1, "classification") :

'train' and 'class' have different lengths

> knn.cv(data = genedata, label = classification, k = 7, p = 12, method = "classification")

Error in knn.cv(data = genedata, label = classification, k = 7, p = 12, :

unused arguments (data = genedata, label = classification, method = "classification")

> res.knn.cv <- knn.cv(data = genedata, label = classification, k = 7, p = 12, method = "classification")

Error in knn.cv(data = genedata, label = classification, k = 7, p = 12, :

unused arguments (data = genedata, label = classification, method = "classification")