

Übungsblatt 8

Übung 8.1 Schreiben Sie ein R-Script, das $n = 50$ normalverteilte Zufallszahlen generiert (Funktion *rnorm*) und die kNN-Dichteschätzung berechnet. Zeichnen Sie zum Vergleich dazu die wahre Wahrscheinlichkeitsdichte, die Sie mit der Funktion *dnorm* erzeugen können.

Hinweis: Eine sortierte Liste der Abstände der Messwerte $x = (x_1, \dots, x_n)$ zum Punkt y erhalten Sie mit

```
dists = sort(abs(y - x));
```

Das Volumen V in der Dichteschätzung $\hat{p}(x) = k/nV$ ergibt sich dann als $V = 2 * dists[k]$.

Übung 8.2 Teilen Sie den in R enthaltenen Iris-Datensatz zufällig auf in Trainingsdaten und Testdaten:

```
ntrain <- 75; ntest <- 150 - ntrain
train <- sample(1:150, ntrain)
iris.train <- iris[train,]; iris.test <- iris[-train,]
```

Bestimmen Sie dann für verschiedene Werte von k die Erkennungsrate des kNN-Classifiers. Dazu können Sie die Funktion *knn* aus dem Package *class* verwenden:

```
classes <- knn(train, test, trainclasses, k)
```

Dabei dürfen *train* und *test* die Klassennamen nicht enthalten, die Sie mit *train[,5]* entfernen können (der Klassenname bei *iris* ist die 5. Spalte) und *trainclasses* darf nur die Klassennamen enthalten. Die korrekten Klassifizierungen können Sie zählen mittels

```
correct <- sum(classes == iris.test[,5])
```

Vergleichen Sie die Erkennungsraten auch mit der Erkennungsrate eines Gaußschen parametrischen Bayes-Classifiers, wie er in dem Script *gaussian.bayes.r* auf der Veranstaltungsseite implementiert ist.