Technische Universität Dortmund
Fakultät Statistik
Lehrstuhl Computergestützte Statistik

Dr. Uwe Ligges
M.Sc. Daniel Horn
B.Sc. Jonas Münch
Moritz Heimbach

Übung zur Vorlesung

Computergestützte Statistik

Wintersemester 2016/2017

Übungsblatt Nr. 13

Abgabe ist Montag der 30.01.2017 an CS-abgabe@statistik.tu-dortmund.de oder Briefkasten 138

Aufgabe 1 **(4 Punkte)**

Implementieren Sie eine k-fache Kreuzvalidierung zur Bestimmung der optimalen Werte der Parameter γ und λ einer regularisierten Diskriminanzanalyse. Optimal meint hierbei eine minimale Fehlklassifikationsrate. Testen Sie Ihre Implementierung auf dem `Ionosphere`-Datensatz aus dem Paket `mlbench`.

Hinweis: Die Implementierung einer `rda` finden Sie im R-Paket `klaR`.

Aufgabe 2 **(4 Punkte)**

In der Vorlesung lineare Modelle haben Sie gelernt, die Güte eines linearen Modells mit dem adjustierten R^2 zu messen, um zum Beispiel Variablenselektion durchzuführen. Im einfachsten Fall kann dazu für jede mögliche Menge von Variablen das adjustierte R^2 bestimmt und die Variablenmenge mit dem besten Wert selektiert werden.

a) (1.5 Punkte) Führen Sie die oben beschriebene Variablenselektion für den `swiss`-Datensatz durch. Mit welcher Variablenmenge können Sie das adjustierte R^2 maximieren?

Alternativ kann auch ein Resampling-Verfahren verwendet werden, um die Güte eines Regressionsmodells zu bestimmen. Dazu wird, wie auch bei der Klassifikation, das Modell auf lediglich einem Teil der Daten trainiert, der übrige Teil der Daten wird für die Validierung verwendet. An Stelle der Misklassifikationsrate schaut man sich bei Regressionsproblemen gerne die mittlere, quadratische Abweichung zwischen tatsächlichen und vorhergesagten Werten an.

b) (1.5 Punkte) Implementieren Sie eine Funktion, die die Güte eines linearen Modells mit einem Resampling-Verfahren bestimmt. Verwenden Sie dafür ein sogenanntes *subsampling*. Bei diesem Verfahren wird die Train-and-Test Methode B Mal ausgeführt und das mittlere Gütemaß zurückgegeben.

c) (1 Punkt) Bestimmen Sie mit Ihrer Funktion aus b) die Güte jeder Variablenmenge der `swiss`-Daten. Welche ist nun am besten?

d) (0.5 Punkte) Erklären Sie den konzeptionellen Unterschied zwischen den Vorgehen aus a) und c).

Aufgabe 3 **(4 Punkte)**

Wie Ihnen allen vielleicht noch aus der Werbung bekannt ist, gibt es zwei einander gegenüberliegende Fabriken, die linke bzw. rechte Twix-Schokoriegel herstellen. Nach langen Streits, welche der beiden Fabriken denn nun den besseren Riegel herstellt, wurden Sie als Statistiker zur Klärung angestellt. Tagesaktuell werden Sie aus beiden Fabriken Daten zu der sogenannten *Crunchigkeit* der Schokoriegel erhalten, und jeden Tag müssen Sie an Hand eines Tests entscheiden: Sind die Riegel aus den beiden Fabriken gleich crunchig, oder gibt es Unterschiede?

Bevor Sie Ihr Tagesgeschäft beginnen können, wollen Sie zunächst den optimalen Test finden, um die Verteilungen der Crunchigkeiten zu vergleichen. Dazu hat Twix Ihnen die Daten dreier Tage zur Verfügung gestellt (siehe Material `chrunchigkeit.RData`). Renommierte Testesser haben dabei in aufwendigen Selbstversuchen festgestellt, dass die Crunchigkeit der beiden Riegel sich an diesen Tagen jeweils nicht unterscheidet. Sie müssen diese Daten nun bestmöglich nutzen und denjenigen Test finden, der die beste Power besitzt. Dabei erinnern Sie sich an den lange zurückliegenden Besuch der Computergestützten Statistik und entscheiden sich einen Bootstrap zu verwenden, um so viel wie möglich aus der einen Stichprobe herauszuholen.

a) (1 Punkt) Implementieren Sie eine Funktion `bootstrapTest`, die einen zweiseitigen Test zum Mittelwertvergleich mittels Bootstrapping implementiert. Rückgabe Ihrer Funktion soll der p-Wert des Tests sein.

b) (2 Punkte) Schreiben Sie eine Funktion `bootstrapPower`. Diese erhält als Eingabe 2 numerische Vektoren x und y — die Crunchigkeiten der beiden Riegel — sowie eine Funktion, die einen statistischen Test ausführt. Diese Funktion soll 2 numerische Vektoren als Eingabe bekommen und den zugehörigen p-Wert zurückgeben. Erzeugen Sie sich mittel Bootstrapping je 200 Stichproben aus x und y, verschieben Sie die Stichprobe y um δ und zählen Sie, wie oft der Test die Nullhypothese ablehnt. Führen Sie dies für mehrere sinnvoll gewählt Werte für δ durch.

c) (1 Punkt) Bestimmen Sie mit Ihrer Funktion aus b) jeweils die Power des t-Tests, des Wilcoxon-Tests und des Bootstrap-Tests für die 3 Tagesproduktionen. Welchen Test empfehlen Sie?