



Blatt 12

Aufgabe 12.1 Bei der Beurteilung der Leistungsfähigkeit von Klassifikatoren spielen die Sensitivität und die Spezifität eine entscheidende Rolle. Oftmals werden dabei ROC-Kurven verwendet, um die Ergebnisse der Klassifikatoren an verschiedenen Arbeitspunkten, d.h. bei verschiedenen Sensitivitäten bzw. Spezifitäten, zu vergleichen.

In dieser Aufgabe wollen wir Ergebnisse der beiden Klassifikatoren K_1^{ML} und K_2^{ML} vom vorherigen Aufgabenblatt miteinander vergleichen, die diese auf dem sigma70-Datensatz (Vordergrund und Hintergrund zusammen genommen) erzielen.

- Beschreiben Sie Ihr Vorgehen, um für die beiden Klassifikatoren die ROC-Kurven zu erhalten.
- Plotten Sie für beide Klassifikatoren die ROC-Kurven auf dem gegebenen Datensatz und vergleichen Sie beide.
- Schätzen Sie für beide Kurven die *area under curve* ab.

Aufgabe 12.2 Für einen bekannten Typ T_1 von Transkriptionsfaktorbindestellen gilt, dass die darin enthaltenen Nukleotide statistisch mit gleicher Wahrscheinlichkeit auftreten. In dem Labor, indem Sie nach Abschluss Ihres Studiums angestellt sind, wurden in einem Experiment Transkriptionsfaktorbindestellen unbekanntem Typs identifiziert. Aus den gegebenen Nukleotidsequenzen der TFBSs haben Sie die Parameter eines homogenen Markov-Modells 0. Ordnung geschätzt, die sich wie folgt ergeben haben:

$$p(A) = 0.23 \quad , \quad p(C) = 0.23 \quad , \quad p(G) = 0.25 \quad , \quad p(T) = 0.29$$

Nun möchte der Biologe, mit dem Sie zusammenarbeiten wissen, ob die gefundenen TFBSs möglicherweise vom Typ T_1 sein könnten, oder ob es sich vielleicht eher um einen anderen Typ von TFBSs handelt. Er sagt Ihnen ausserdem, dass er gehört hat, dass zur Lösung dieses Problems wohl oft ein χ^2 -Test Anwendung findet, genaueres weiss er aber nicht.

- Finden Sie heraus, was der χ^2 -Test ist und beschreiben Sie in wenigen Sätzen das Prinzip dieses Tests. Geben Sie die Quellen Ihrer Informationen an.
- Beschreiben Sie die Vorgehensweise, um mit Hilfe des Tests Ihr TFBS-Problem zu lösen. Führen Sie den Test mit den gegebenen Daten durch. Welche Antwort können Sie Ihrem Kollegen anschließend geben?