



Einführung in die Bioinformatik

SoSe 2013

Übungsblatt Nr. 10

Lernen mit Kernen

1. In dieser Aufgabe erkunden wir verschiedene Vorgehensweisen, um aus bekannten Kernen neue Kerne abzuleiten.

Wir nehmen an, dass $x, x' \in \mathcal{X}$ und definieren $k_1(x, x')$ und $k_2(x, x')$ als Kerne über $\mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$.

Beweisen Sie folgende Aussagen:

- (a) **Skalieren:** Gegeben sei ein Skalar $a \in \mathbb{R}^+$. Beweisen Sie, dass $k(x, x') := ak_1(x, x')$ ein Kern ist.
- (b) **Summe:** Zeigen Sie, dass $k(x, x') := k_1(x, x') + k_2(x, x')$ ein Kern ist.
- (c) **Produkt:** Zeigen Sie, dass $k(x, x') := k_1(x, x')k_2(x, x')$ ein Kern ist.

Hinweis: Oft kann man zeigen, dass eine Funktion k ein Kern ist, indem man die entsprechende Abbildung ϕ in den Feature Raum angibt.

2. In dieser Aufgabe untersuchen wir String-Kerne, die die Häufigkeiten von k -meren (Substrings der Länge k) in zwei Strings, z.B. zwei DNA-Sequenzen, miteinander vergleichen.
 - (a) Wie viele 2-mere gibt es im DNA-Alphabet $\{A, C, T, G\}$? Geben Sie die gesamte Liste an.
 - (b) Definieren Sie einen Kern, der die 2-mere in 2 Strings vergleicht.
 - (c) Wie lautet die allgemeine Formel, um die Menge aller k -mere im DNA-Alphabet zu bestimmen für ein fixes k ? Berechnen Sie diese Werte für k von 3 bis 9.
 - (d) Aus wie vielen Termen besteht ein Kern, der 9-mere in zwei Strings vergleicht?