



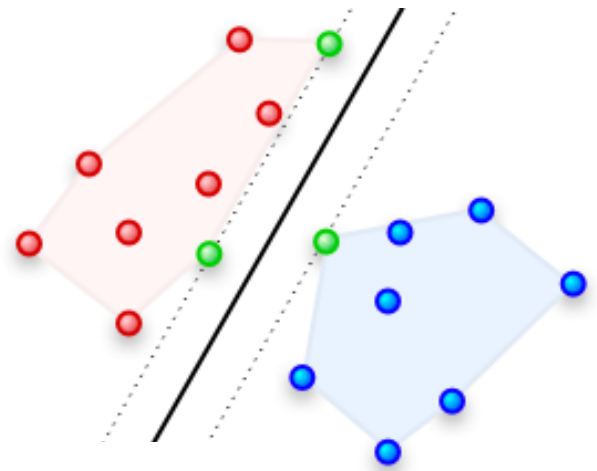
Feature-Kodierung von SNP-Daten für Support Vector Machines

In Genome Wide Association Studies (GWAS) zur Untersuchung von Krankheiten werden große Datenmengen produziert, die ohne geeignete Programme nicht effizient verarbeitet werden können. Maschinelle Lernverfahren, wie z.B. Support Vector Machines (SVM), bieten erprobte Verfahren zur Erkennung von Mustern in solch großen Datenmengen. Dennoch sind geeignete Vorverarbeitungsschritte notwendig um brauchbare Ergebnisse zu liefern. So gibt es verschiedene Möglichkeiten zur Kodierung der SNP-Daten, die verschiedene Risikomodelle implizieren.

In dieser Bachelorarbeit sollen verschiedene Arten der Feature-Kodierung für SNP-Daten für die Verwendung in Support Vector Machines implementiert und bezüglich der Performanz der daraus resultierenden Ergebnisse verglichen werden. Die Implementierung erfolgt in Java und soll auf bereits vorhandenen Tools basieren (z.B. LibSVM).

Anforderungen:

Gute Kenntnisse in Java werden vorausgesetzt.
Erfahrungen im Maschinellen Lernen (insbesondere SVMs) sind hilfreich



Kontakt

Florian Mittag
Sand 1, Raum A306
Tel. (07071) 29-78985
florian.mittag@uni-tuebingen.de