

## Übung zur Vorlesung **Maschinelles Lernen**

Wintersemester 2008/2009  
Blatt 13

### **Aufgabe**

10 Punkte

In der letzten Vorlesung ging es unter anderem um das Thema *Sampling*. Dabei wurden unterschiedliche Sampling-Strategien vorgestellt (*generic sequential sampling, stratified sampling*).

Innerhalb von RapidMiner erlaubt z.B. die Kreuzvalidierung verschiedene Sampling-Methoden.

1. Wiederholen Sie zwei Ihrer früheren Experimente mit Kreuzvalidierungen (z.B. Mehrklassen-Problem mit der SVM (Blatt 8), Text-Klassifizierung (Blatt 9)) und verwenden Sie dabei jeweils verschiedene Sampling-Methoden. Sie können dazu natürlich auch gerne ein neues Klassifizierungsproblem erstellen.
2. Wie haben sich die verschiedenen Sampling-Strategien auf Ihre Ergebnisse ausgewirkt? Stellen Sie Ihre Ergebnisse *präsentationsfähig*<sup>1</sup> dar.

**Hinweis:** Die Erstellung eigener Daten ist natürlich etwas aufwendiger, es wäre aber schön, wenn wir bei der Besprechung der Abgabe verschiedene Experimente hätten. Daher appelliere ich an dieser Stelle einfach mal an Eure Kreativität und hoffe, dass nicht jeder die Experimente aus Blatt 8 und Blatt 9 untersucht.

Insbesondere wäre es ja auch interessant, wie sich die Sampling Strategien bei der Kreuzvalidierung auf Experimente mit Euren eigenen Klassifikationsoperatoren auswirken.

---

<sup>1</sup>Es wäre schön, wenn jede Gruppe bei der Besprechung die eigenen Ergebnisse in ansprechender Weise präsentieren könnte.