

Lehrveranstaltung “Maschinelles Lernen 2 – Fortgeschrittene Verfahren“

Übungsblatt 1

Semi-Überwachtes und Aktives Lernen

Anmerkungen:

- Lösen Sie anhand des in der Vorlesung vermittelten Stoffes die Aufgaben
- Nutzen Sie wenn nötig die Programme aus ILIAS
- Fragen und Lösungen werden in den Übungen besprochen

Aufgabe 1: Semi Supervised Learning (SSL)

- (a) Beschreiben Sie kurz überwachtes und semi-überwachtes Lernen. Legen Sie hierbei den Fokus auf die Unterschiede:

Unüberwacht:

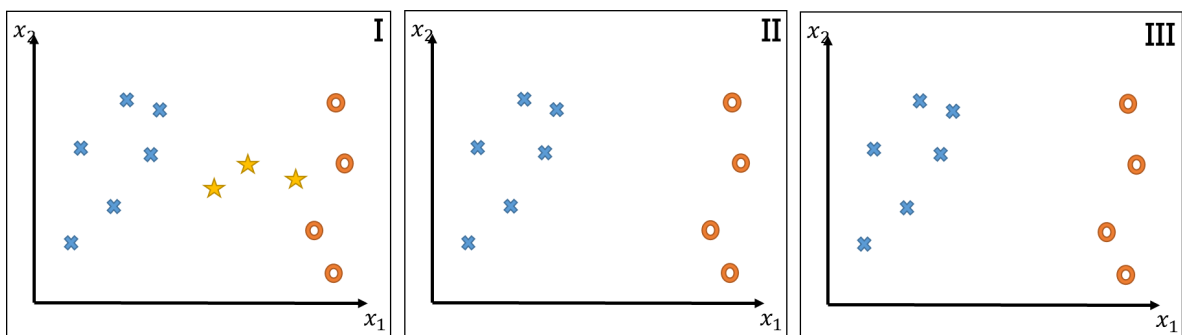
Semi-überwacht:

- (b) Für das semi-überwachte Lernen werden einige Annahmen getroffen. Erläutern Sie die Dichte-Annahme. Welche Folgerungen können daraus geschlossen werden?

(c) Erläutern Sie den Unterschied zwischen Induktion und Transduktion:

(d) In der folgenden Grafik finden Sie innerhalb des Schaubildes I den Ausgangspunkt für einen Self-Learning Algorithmus. X und O stellen dabei bereits gelabelte Daten dar. Die Sterne stehen für ungelabelte Trainingsdaten.

- i. Zeichnen Sie nun die in Schritt I ungelabelten Datenpunkte mit den jeweiligen Labeln in den Schritten II und III ein. Der Algorithmus berücksichtigt pro Schritt nur einen ungelabelten Datenpunkt:



- ii. Nach welchem Kriterium werden die Datenpunkte ausgewählt?

- iii. Welche Gefahr besteht beim Self-Learning? Welchen Lösungsansatz kennen Sie?

Aufgabe 2: Aktives Lernen

- (a) Welche drei verschiedenen anwendungsbezogenen Ausprägungen des aktiven Lernens gibt es? Benennen und erläutern Sie diese kurz und prägnant:

- (b) Gegeben sei das Lernproblem der bildbasierten Klassifikation von Verkehrszeichen. Ein menschlicher Experte erhält von der aktiven Lernmaschine eine Anfrage bzgl. eines zu klassifizierenden, künstlich synthetisierten Lernbeispiels. Was ist die Herausforderung für den Experten gegeben der angefragten Daten?

- (c) Welche drei Unsicherheitsmaße können zur Datenauswahl beim aktiven Lernen herangezogen werden?

- (d) Füllen Sie zur Erläuterung des aktiven Lernens im Version Space folgende Lücken aus. Beim aktiven Lernen im Version Space werden die ungelabelten Daten gesucht, die

Beim SVM-„Simple Margin“-Kriterium werden die Label der Datenpunkte angefragt, deren

- (e) Geben Sie das mathematische Kriterium für die Auswahl des anzufragenden Datenpunkts x gemäß MaxMin-Margin und Ratio-Margin an.

Aufgabe 3: Programmbeispiel

Im ILIAS finden Sie das zip-Archiv „s3vm_toy.zip“. Nach dem Entpacken finden Sie in dem Ordner „s3vm_toy“ das Java-Applet s3vm_toy. Dieses basiert auf der häufig verwendeten LIBSVM¹ und wurde für das semi-überwachte Lernen dahingehend angepasst, dass als „weiß“ eingegebene Daten als ungelabelt betrachtet werden. Sofern ungelabelte Daten vorhanden sind, werden beim Lernen (aktiviert durch en Button „s3vm“) nach dem Prinzip des Selbstlernens iterativ für jede Klasse je ein neuer Datenpunkt der ungelabelten Daten gelabelt. Ist die Maschine so eingestellt, dass eine probabilistische SVM verwendet wird (Parameter „-b 1“) wird jeweils der aktuell konfidenteste Vertreter gewählt, sonst (Parameter „-b 0“) wird jeweils der nach Eingabe nächste, entsprechend klassifizierte Datenpunkt gelabelt. Sie werden feststellen, dass dies Auswirkungen hat.

Sie finden auf der offiziellen Seite der LIBSVM eine vollständige Parameterbeschreibung, die weiterhin gültig ist. Zur einfachen Verwendung stehen im Archiv weiterhin einige Datensätze bereit. Sie können die App auf der Kommandozeile mit „java s3vm_toy“ starten.

(a) Stellen Sie in der App die Aufgabe 1.c nach. Sie können dazu auch den Datensatz „example_simple“ verwenden.

- i. Vergleichen Sie die Ergebnisse.

- ii. Starten Sie die Berechnung wiederholt. Vergleichen Sie die Ergebnisse. Welche Schlüsse ziehen Sie? Wie erklären Sie die unterschiedlichen Ergebnisse.

- iii. Variieren Sie die Parameter. Z.B. den Kerntyp über den Parameter „-t 0“ für lineare Kernel und „-t 2“ für RBF – Kernel oder über den Parameter „-b 1“ für die probabilistische SVM bzw. „-b 0“ nicht probabilistisch. Was passiert dabei?

¹ Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

- iv. Testen Sie die weiteren Datensätze `example1` bis `example5` mit unterschiedlichen Parametern. Erklären Sie die Ergebnisse.

- (b) Laden Sie den Datensatz `example_2_moons_plain`. Ziel ist es den Eingaberaum in Form zweier „Halbmonde“ zu trennen.

- a. Überlegen Sie, wie Sie „aktiv“ durch Hinzufügen weniger gelabelten Daten den Lernprozess starten.

- b. Testen Sie das Lernen auf den Datensätzen `example_2_moons_1`, `example_2_moons_2`, `example_2_moons_3`. Was beobachten Sie? Erklären Sie die Ergebnisse.

- (c) Wie könnten Sie das Verhalten des SSL verbessern um das Lernen zu beschleunigen?