

Arbeitsbericht

Projekt: LDA & SVM

Autor: Henning Stein

Das Projekt begann damit, dass wir ausmachten uns über die Weihnachtstferien in den Stoff einzulesen und uns gegen Ende zu treffen um erste Dinge auszumachen. Bei diesem Treffen wurden jedem zufällig ein Themengebiet zugeordnet:

- Niklas: LDA und QDA
- Bennedikt: SVM
- Henning: PDA
- Daniel: FDA

Zusätzlich war die grobe Idee innerhalb von ca. zwei Wochen eine halbwegs funktionierende Implementierung der gegebenen Themengebiete zu haben. Dabei stellte sich jedoch schnell heraus, dass die gegebene Quelle für manche Gebiete (hauptsächlich SVM und FDA) nur mäßig gut ist. Da Niklas recht schnell eine sehr gut funktionierende Version von LDA und QDA lieferte und sich PDA sehr stark daran orientiert konnte ich einiges des gegebenen Codes übernehmen womit auch nach schon ca. einer Woche das Grundgerüst für PDA stand. Damit blieb noch die Basisexpansion übrig, bei der ich den anfänglichen Versuch beliebige polynomiale Expansionen zu wählen bald aufgab und mich auf 2. und 3. Grad beschränkte. Für den letzten Punkt (die Projektion auf einen zweidimensionalen Unterraum) konnte ich, nachdem mir Christof eine weitere Quelle geschickt hat, Niklas Funktion verwenden, der schon vorher auf das gleiche Problem stieß.

Als in der Vorlesung shiny behandelt wurde fasste ich recht schnell den Beschluss, dass das Projekt davon auch profitieren würde. Einerseits fand ich eine graphische Oberfläche sehr ansprechend, andererseits hatte ich das Gefühl, dass ich, da ich viel von Niklas übernahm, zu wenig hatte und nahm mir damit eine Zusatzaufgabe. Die grundsätzliche Implementierung war dann schon am selben Abend fertig und von da an galt es auf den Fortschritt der anderen zu reagieren und die App zu updaten.

Anfangs hatte ich noch die Idee die Möglichkeit einzubauen einn eigenen Testsatz zu erstellen indem man die Punkte in einem Plot einsetzt. Als ich dann soweit war, dass ich die Punkte einsetzen konnte, entschloss ich die Idee zu verwerfen, da es zu aufwendig ist einen Datensatz manuell zu erstellen und es nur eine grundsätzlich nutzlose Funktion ist, deren einziger Nutzen ist schön auszusehen.

In der letzten Woche, nachdem alle Klausuren geschrieben waren und man sich vollständig auf das Projekt konzentrieren konnte, setzte sich Niklas daran das Projekt umzustrukturieren und auf das R6 Objekt zu basieren. Glücklicherweise ergaben sich damit für mich keine weiteren Schwierigkeiten, da PDA sehr ähnlich zu LDA und QDA funktioniert und ich in Shiny nur die Aufrufe der anderen übernehmen musste. Also gab es noch die Aufgabe shiny ein bisschen nützlicher und schöner zu machen. Dafür fügte ich zuerst die Option hinzu Plots als Bilder zu speichern. Dabei stellte sich die Schwierigkeit, dass sich eine Shiny-App (soweit ich weiß) strukturell von einer normalen Funktion unterscheidet und damit auch die scopes der Variablen nicht funktionieren wie in normalen Funktionen. Um das zu lösen erstellt die Shiny-App nun beim Aufruf eine eigene Umgebung in der die Objekte gespeichert werden und bei er es kein Problem gibt an einem späteren Zeitpunkt aus der App darauf zuzugreifen. Als nächstes fügte ich die Funktion hinzu Testobjekt zu laden um die App für mehr zu nutzen als nur zufällige Plots. Schließlich gab es noch einige kosmetische Aufgaben wie die Funktion, dass manche Optionen erst dann auftauchen, nachdem andere ausgewählt wurden (z.B. Kernel als Reaktion auf die Wahl von SVM). Glücklicherweise stellt shiny mit conditional Panels bereits eine Funktion dafür.

Ich entschied mich dafür, da ich keine neue Programmieraufgabe so kurz vor dem Ende anfangen wollte, die Vignette zu schreiben. Anfangs war ich mir unsicher was ich schreiben sollte und begann etwas ausführlicher zu. Dabei stellte sich aber heraus, dass die Vignette mit dem Einfügen von plots sehr schnell sehr groß wird und nach kurzer Zeit war ich bei 15 Seiten angelangt, die aber zu mehr als der Hälfte aus plots bestehen. Da das Paket aber vorallem mit schönen plots glänzen kann, lässt sich darauf aber nur schwerlich verzichten. Außerdem hielt ich es für sinnvoll und notwendig zu jedem der Klassifizierungsalgorithmen eine plot als Beispiel zu zeigen. Für das motivierende Beispiel am Anfang der Vignette habe ich einen eigenen Datensatz Rockets erstellt, der sich an Sandras Übungsaufgabe für Sylvester aus der „Einführung in die Wahrscheinlichkeitstheorie und Statistik“ orientiert. Die Tatsache, dass die Lautstärke der Raketen mit deren Flughöhe korreliert bleibt hierbei außer Acht gelassen. Der Grund dafür, dass der Datensatz selbst erstellt wurde, war, dass wir Probleme hatten geeignete real data im Internet zu finden aber für die Präsentation des Pakets nicht auf ein Anwendungsbeispiel verzichten wollten.

Die Vignette beinhaltet zusätzlich zur Funktionsweise des Pakets eine kurze Beschreibung der Funktionsweise und Idee der Algorithmen. Da sich nicht viel mehr an Inhalt schreiben ließe ohne sehr viel tiefer in die Mathematik einzugehen, was niemand will (siehe Hastie et al. „The Elements of Statistical Learning“ (2009) Seite 420), ist die Vignette recht nah an einer educational vignette zu diesem Thema und kann im Zweifel als solche betrachtet werden.