

# Bericht

## Niklas Wünstel

Bei der Themenauswahl waren wir uns relativ schnell einig. Wir haben alle Themen kurz überflogen und uns danach auch keine weiteren Gedanken gemacht. Als uns LDA & SVM zugewiesen wurde, haben wir uns über die Weihnachtsfeiertage erstmals wirklich durchgelesen, worum es überhaupt geht. Es wurde mir schnell klar, dass das Projekt anspruchsvoller sein wird, als ich gehofft hatte. Einiges zum Thema war uns nach den Ferien immer noch unklar. Weil uns klar war, dass wir Ende Januar und Anfang Februar keine Zeit zum Programmieren haben werden wegen den anstehenden Klausuren, hatten wir es deshalb aufgegeben, alle alles zu verstehen. Wir haben die einzelnen Kapitel in vier Teile aufgeteilt und zufällig zugewiesen. Dabei hatte ich das Glück, dass ich den einfachsten Teil bekommen hatte (LDA & QDA).

Ich musste nicht mehr viel lesen, weil ich das meiste dazu schon in den Ferien verstanden hatte und konnte deshalb auch als Erster anfangen zu programmieren. Ich habe Bitbucket eingerichtet und einfach mal angefangen. Alle wichtigen Schätzer (Mittelwert, Varianz etc.) für LDA & QDA und auch eine einfache Variante der jetzigen `make_test` Funktion hatte ich schnell geschrieben. Das war für alle nützlich, um Code zu testen. Eine Woche nach den Weihnachtsferien konnte ich schon einfache selbsterstellte Test-Datensätze via LDA & QDA klassifizieren und mit ggplot plotten. Alle Funktionen waren relativ einfach und das Programmieren hat mir Spaß gemacht, obwohl ich schon auch mehrmals bis spät in die Nacht programmiert hatte.

Henning konnte viele meiner Funktionen sehr gut übernehmen und ist auch schnell vorangekommen. Benni hatte mit SVM das schwerste Thema erwischt und konnte erst nach einem Gespräch mit Christoph eine funktionierende Klassifizierungsfunktion erstellen. Daniel konnte mit der Erklärung von FDA aus „The Elements of Statistical Learning“ nicht viel anfangen und durfte sich eine andere Klassifizierungsmethode aussuchen. Das Buch war eine wirklich umständliche Quelle. Man musste viele Kapitel lesen, um Dinge ausreichend gut verstehen zu können, dass man sie in R implementieren konnte, vor allem, wenn man von der Thematik noch nicht viel gehört hatte.

Allerdings hatten wir ca. drei Wochen nach den Weihnachtsferien viel funktionierenden Code. Die Dateien wurden immer länger und wir haben erstmals angefangen Funktionen sinnvoll in mehrerer Dateien aufzuteilen, um alles übersichtlicher zu machen. Viele Dinge wurden in jeder Klassifizierungsmethode neu berechnet und generell hatten wir viel

Redundanz in LDA QDA PDA und RDA. Ich hatte deshalb zu dem Zeitpunkt den Vorschlag gemacht, OOP zu benutzen um Wiederholungen zu vermeiden und die Verwendung von unseren Funktionen leichter zu machen. Umgesetzt habe ich das allerdings erst nach den Klausuren. Der einzige der trotz der Klausuren noch regelmäßig an seinem Code geschrieben hatte war Benni.

Nach der Klausurenphase haben wir uns erst mal einen Überblick verschafft und eine ToDo Liste erstellt. Ich habe dann die R6 Klasse ‚data\_set‘ erstellt und die meisten Funktionen, die wir schon geschrieben hatten, damit kompatibel gemacht. Im Nachhinein hätte man das direkt am Anfang so umsetzen müssen - dass hätte viel Verwirrung verhindert und Arbeit gespart. Allerdings hatten wir am Anfang einfach noch keinen richtigen Überblick über das Thema und keiner wusste wirklich wohin alles führen wird. Jedenfalls sind wir alle der Meinung, dass diese Umstellung vieles verbessert hat. Ich konnte über die Klasse viele Checks laufen lassen, Fehler vermeiden, und mehr Kontrolle über die Daten bekommen.

Ein weiterer größerer Störfaktor war, dass wir alle viel zu viel über Bitbucket gearbeitet haben. Jeder hat mehrmals täglich teils unfertigen Code gepusht und weil wir alle unseren Code direkt im Paket getestet haben, hatte man nach dem pullen meistens irgendwelche Fehlermeldungen von denen man keine Ahnung hatte. Außerdem war Mergen das reinste Chaos. Daniel hat z.B. auch lange Zeit mit veraltetem Code gearbeitet, weil wir alten Code nicht regelmäßig gelöscht haben. Alles in allem hatten wir wirklich viele Probleme mit Version Control würde ich sagen.

Wir haben uns zwar auch im Januar öfters getroffen und ausgetauscht, aber alles zusammengesetzt haben wir erst in der Woche vor den Vorträgen. Vorher hatte ich mir einigen Code garnicht angeschaut. Wir haben öfters Übungsräume in der Altstadtbibliothek reserviert und einzelne Problem die noch übrig waren, gemeinsam gelöst. Restliche kleinere Aufgaben wie z.B. auskommentieren oder die Vignette schreiben haben wir erst ganz zum Schluss gemacht. Unseren Vortrag haben wir am Tag vorher zusammengebastelt. Markdown bietet dafür ja wirklich gute Möglichkeiten. Trotz, dass wir meiner Meinung nach viel vor den Klausuren gemacht hatten, war die letzte Woche nochmal wirklich stressig. Wir haben alle sicher weit über 50 Stunden Arbeit in dieses Paket gesteckt. Dennoch hat es mir wirklich viel Spaß gemacht und ich finde auch, dass ich viele Erfahrungen im gemeinsamen Programmieren gesammelt habe.

Funktionen die ich geschrieben habe:

Calc\_error.R: Alles,

Classifier\_funs.R: von targets bis inclusive QDA

Estimators.R: alles bis auf sigma\_bet\_exp und small\_sigma\_est

oop.R: Alles,

Plot\_functions.R: Alles,

Test.R: Alles

