

Statistical Inference for Data Science with Innovative Teaching Methods

Projektverantwortliche / Projektverantwortlicher

PD Dr. Sigve Haug

Institut

Mathematisches Institut

Projektlaufzeit

HS 2017/FS 2018

Abstract

Die digitale Revolution und die ihr zugrundeliegende technische Entwicklung lassen die Menge der empirischen Daten immens vergrössern. Damit verbessern sich die Möglichkeiten rationaler Schlüsse und Entscheide, nicht nur in der Forschung, sondern auch auf vielen gesellschaftlichen und wirtschaftlichen Ebenen. Die Entfaltung des positiven Potentials bedingt aber theoretische und praktische Kenntnisse in *Data Science* und der Bedarf an Angeboten steigt in vielen Fachrichtungen.

In diesem Projekt wird als Teil eines grösseren Portfolios von praktischen *Data Science* Angeboten ein *Statistical Inference* Kurs mit innovativen Lehrmethoden entwickelt. Meine Daten sind da, wie extrahiere ich korrektes Wissen aus ihnen ? Welche Werkzeuge gibt es, z.B. R, MatLab, Python ? Wie benütze ich sie ? Welche Rechen- und Lagerungsinfrastrukturen stehen mir zur Verfügung, z.B. das lokale Linux Cluster, das nationale SWITCHengines oder eine der schnellsten Maschinen der Welt in Lugano ? Das Ziel ist möglichst vielen Fachrichtungen und Niveaus, von Bachelorstudierenden bis Assistierenden, bei solchen Fragen zu dienen. Dabei wird Fokus nicht bei formaler Vollständigkeit, sondern beim praktischen Umgang mit vorhandenen Werkzeugen und bei korrekter Interpretation und Darstellung gesetzt.

Mit Hilfe alternativer und innovativer Präsenzmodelle, ausgedehnter Integration von *e-Learning* Techniken und Plattformen und ebenbürtiger Evaluation wird die Herausforderung der unterschiedlichen Stundenpläne und Vorkompetenzen angegangen. Wohldefinierte Lernziele für jede einzelne Präsenzveranstaltung, Zugang zu Lernmaterialien unterschiedlicher Stufe und Teilassessments unterstützen das gemischte Lernen (*blended learning*) und die optimale Vorbereitung auf die Präsenzveranstaltungen, in denen der Stoff diskutiert und nicht vorgelesen wird (*flipped classroom*). Um die kollegiale und berufsnaher Umgebung zu unterstreichen oder üben, werden Teilnehmende zum Teil sich selbst evaluieren (*peer evaluation and peer assesement*).

Fazit

In diesem Projekt haben wir in einer massgeschneiderten Veranstaltung “Statistical Data Analysis with Python, R and Prism” (2 ECTS), https://ilias.unibe.ch/goto_ilias3_unibe_crs_1274865.html, browserbasierte (Jupyter) Notebooks mit dem *inverted classroom* Konzept verbunden. Das interaktive Ausführungs- und Darstellungsvermögen der Notebooks verstärkt die potentiellen Vorzüge des sogenannten umgedrehten Unterrichtes.

Sorgfältig aufgearbeitete Notebooks, von den Lehrkräften erstellt, bieten viele Vernetzungsvorteile des *World Wide Webs* und eine interaktive Programmierungsumgebung mit unmittelbarer Ausführung und Darstellung, alles in einem Webbrowser. Die Notebooks können entweder lokal auf eigenem Gerät oder als Service über eine Internetverbindung bearbeitet werden. Die Interaktivität stimuliert die selbständige Erarbeitung des Stoffes vor der Präsenzveranstaltung mit einer zusätzlichen Komponente des Ausprobierens und Nachvollziehens. Die Möglichkeit mit Parametern in Beispielen zu spielen und unmittelbar die Effekte in den graphischen und textuellen Darstellungen zu beobachten, fördert das Verständnis. Diese Annahme hat sich im Projekt bewährt.

Wir hatten in diesem Projekt Präsenzzeiten von 90 Minuten für die Diskussion des Stoffes der vorher bearbeiteten Notebooks. Am Tag vor der Präsenzveranstaltung war es Pflicht, d.h. Teil des *Assessments*, Fragen zum Stoff einzureichen (via Google Forms). Diese Fragen wurden in der Präsenzzeit diskutiert. Die Teilnehmenden fanden 90 Minuten als adäquat für die Stoffmenge der Notebooks, mit denen ein erwarteter Aufwand von etwa drei Stunden verbunden war. Sie bevorzugten auch eine Mischung aus Selbstbeantwortung in Gruppen und Dozentenbeantwortung der Fragen.

Die Teilnehmenden beurteilten ihre (Team-) Projekte, die als Poster vorgestellt wurden, mit einem vorgegeben Raster (Peer-Assessment). Der Dozent fand die Benotung etwas zu positiv. Eine weitere Quantitativisierung des Rasters könnte dies entgegenwirken.

Die nächste Durchführung des Kurses wird die gleichen didaktischen Elemente beinhalten. Die Notebooks sollen weiter verbessert und etwas mehr Zeit in die Strukturierung der Diskussionsrunden investiert werden. Reine Beantwortung und Erläuterung der Fragen soll möglicherweise mit konkreten Aufgaben in der Präsenzzeit ergänzt werden.

Der sogenannte umgedrehte Unterricht in Kombination mit Jupyter Notebooks ist ein hervorragendes Format für Veranstaltungen, in denen die Verarbeitung von Daten und Zahlen Teil der Lernziele ist. Es fängt mit der Arbeit mathematischer und statistischer Bibliotheken und hört nicht mit der Mustererkennung und Diagnose in Bildern der Psychologie und Medizin auf. Abgesehen von der hier beschriebenen Veranstaltung, “Statistical Data Analysis with Python, R and Prism”, wird das Format nun am Mathematischen Institut im CAS Applied Data Science und in einigen *Digital Skills Training* Kursen eingesetzt.

Im Rahmen der Digitalisierungsstrategie der Universität Bern ist es mittelfristig zu erwarten, dass Lernangebote im Bereich Datenanalyse flächendeckender angeboten werden müssen. Das in diesem Projekt ausprobierte Format wird für viele dieser Angebote sicherlich eine Option sein. Ab Januar 2019 steht ein Crashkurs über Jupyter Notebooks und deren Anwendung im Digital Skills Training Programm des Science IT Support zur Verfügung (scits.unibe.ch/courses).