Repetitorium, 12. Mai 2017

Dayyan Smith

Signifikanztests

Für Machine Learning Experimente werden die vorhandenen Daten meist dreigeteilt. Das Training Set wird benutzt, um Parameter automatisch zu lernen, das Development Set wird benutzt um Design Entscheidungen zu treffen und mit dem Test Set kann man einen Schätzwert für Performance auf neuen Daten berechnen. Diese drei Sets dürfen sich nicht überschneiden. Wenn ein System B gemäß eines Evaluationsmaß (z. B. Accuracy) besser ist als ein System A, wie sicher ist System B dann auch auf neuen Daten besser? Mit Signifikantests kann ein Wert berechnet werden, der angibt wie wahrscheinlich es ist dass der Unterschied in der Performanz zweier Systeme zufällig ist. Oft möchte man testen, ob es überhaupt einen Unterschied zwischen zwei Systemen gibt. In dem Fall wäre die Nullhypothese: Es gibt keinen Unterschied zwischen System A und System B. Weiterhin gibt es eine Teststatistik t(X, Y) wobei X und Y die Outputs zweier Systeme sind, die den Unterschied der Evaluationsmaße berechnet. Dann sucht man eine Verteilung über t(X, Y) unter der Nullhypothese. Nun kann man sehen wo der beobachtete Unterschied zwischen System A und System B in dieser Distribution liegt. Wenn extremere Werte wahrscheinlich sind, so behält man die Nullhypothese. Sind sie es nicht, so wird die Nullhypothese verworfen.