

Projektarbeit: Anomalieerkennung mittels GP-LVMs

Generative Modelle beschreiben eine Klasse von Machine Learning (ML) Modellen die "quasi aus dem nichts" Daten generieren können. Beliebte Beispiele sind das Erzeugen von Deepfakes in Form von Videos und Bildern, sowie Chat-Systeme die auf die Eingaben von Benutzern mit sinnvollem Text reagieren. (Siehe Figure 1 Links/Mitte)

Die Klasse von generativen Modellen die uns in diesem Projekt interessieren sind die sogenannten **Gaussian Process Latent Variable Models (GP-LVMs)**. Diese Modelle lernen die Abbildung von einem **latenten Raum**, ein zugrunde liegender "virtueller" Raum mit viel Struktur (vgl. Figure 1 Rechts), zurück in unseren echten Raum in dem die Daten liegen.

Diese Modelle sind zwar sehr Dateneffizient, müssen aber den latenten Raum in einem aufwendigen Prozess aus den echten Daten lernen. Daraus folgt: für jeden neuen Datenpunkt muss der gesamte Lernprozess wiederholt werden!

Das soll in dieser Arbeit verbessert werden! Sie sollen einen gewöhnlichen **Gaußprozess mit einem GP-LVM verbinden** um einfacher mit neuen Daten umgehen zu können. Das passiert durch das Lernen einer Abbildung des Datenraums in den latenten Raum. Zusätzlich kann untersucht werden, ob beiden Modelle gleichzeitig trainiert werden können, um das Training zu verbessern.

Das fertige Modell kann anschließend z.B. in der Anomalieerkennung eingesetzt werden, um Probleme in neuen Datenpunkten zu erkennen.

In dieser Arbeit werden Sie mit wissenschaftlicher Literatur und einer etablierten open source Bibliothek im Bereich des ML (GPyTorch) arbeiten um ein simples neuartiges Modell zu entwickeln. Dadurch werden Sie Hands-On Erfahrungen mit Gaußprozessen sowie GP-LVMs sammeln und die Theorie verstehen.

Gruppengröße: 3-5 Studierende mit Vorkenntnissen im maschinellen Lernen

Betreuung: M.Sc. Andreas Besginow und Prof. Dr. Lange-Hegermann.

Sollten Sie bereits Fragen zum vorgestellten Thema haben, können Sie sich gerne an folgende E-Mail Adresse wenden: andreas.besginow@th-owl.de

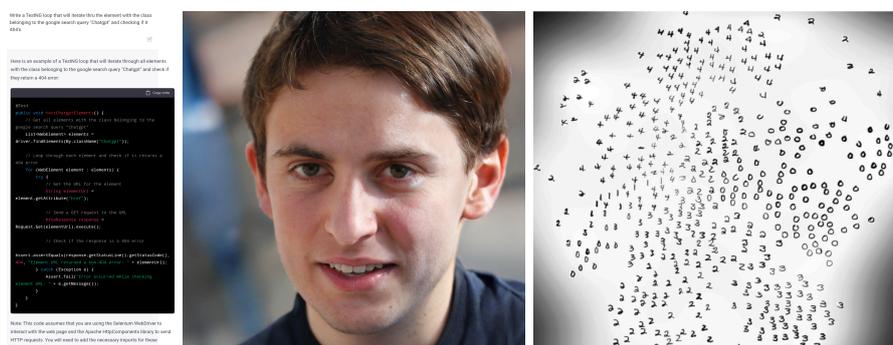


Figure 1: (Links) Generiert durch chatGPT. (Mitte) Generiert durch www.thispersondoesnotexist.com. (Recht) Aus Lawrence, Neil. "Gaussian process latent variable models for visualisation of high dimensional data."