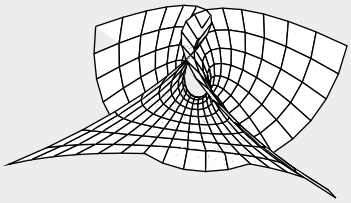

Oberseminar Mathematische Strömungsmechanik

Institut für Mathematik der Julius-Maximilians-Universität Würzburg

Leon Bungert

Universität Würzburg

The Geometry of Adversarial Machine Learning

Abstract:

It is well–known that despite their aptness for complicated tasks like image classification, modern neural networks are prone to insusceptible input perturbations (a.k.a. adversarial attacks) which can lead to severe misclassifications. Adversarial training is a state–of-the-art method to train classifiers which are more robust against these adversarial attacks. The method features minimization of a robust risk and has interpretations as game-?theoretic problem, distributionally robust optimization problem, dual of an optimal transport problem, or nonlocal geometric regularization problem. In this talk I will focus on the last interpretation which allows for the application of tools from calculus of variations and geometric measure theory to study existence, regularity, and asymptotic behavior of minimizers. In particular, I will show that adversarial training of binary agnostic classifiers is equivalent to a nonlocal and weighted perimeter regularization of the decision boundary. Furthermore, I will show Gamma–convergence of this perimeter to a local anisotropic perimeter as the strength of the adversary tends to zero, thereby establishing an asymptotic regularization effect of adversarial training.

room 40.03.003 (Emil Fischer Str. 40)

Thursday, Oct. 19 at 12:30 pm

Zu diesem Vortrag sind Sie herzlich eingeladen.

gez. Christian Klingenberg