Adversarial training protects the **non-robust** features.
A **trade-off** emerges if those features are **useful**.

A High Dimensional Statistical Model for Adversarial Training: Geometry and Trade-Offs

Kasimir Tanner
Matteo Vilucchio
Bruno Loureiro
Florent Krzakala

EPFL