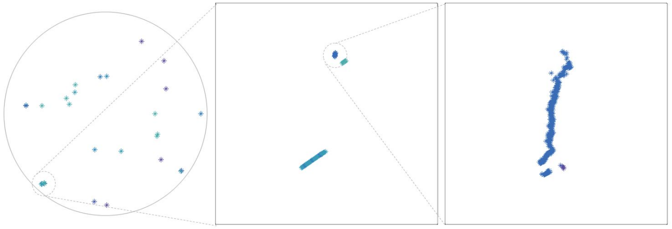
Defect!D

Aruna Gunda ⁛ Sreesha Nath ⸬
Tian Zhang ⸌ Nathan Kot ⸪



In any form of business it is ideal for the company and the world that operations are as efficient as possible. Sometimes this means admitting you've made a mistake as early as possible.

The fly-height of the read-write head in modern hard-drives is in the tens of angstroms. Disk spin-speeds can reach RPMs of 10k. Can you imagine hitting a "bump" larger than your vehicle while traveling at 50,000mph?

$$P(day\ ruined \mid 50kmph\ collision) = 1$$

In addition to roughness the sputtering process does not always go well. Contamination mitigation does not always go well. The head does not always function well. etc. etc.

The earlier one can identify such defects the less we waste.

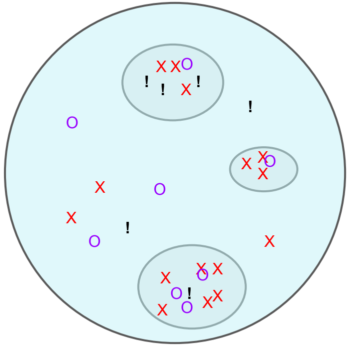
Background

- Hard-drive disk manufacture is the product of a multi-step process
- There are opportunities within said process to attempt to quantify a disk's quality with respect to a myriad of metrics
- There is balance in pursuing this, with confidence in one's identification often traded for throughput
- This trade-off, paired with the fact identification requires angstrom-level resolution, means one's confidence can *really* suffer
- The goal of this project is to attempt to pair knowledge of ML with disparate, sparse feature sets to make confident predictions as to whether or not they reveal a true defect
- The feature sets are supported by a label set produced after the drive has been fully assembled, where confidence in defect identification is quite high but, alas, it is too late

Initial approach

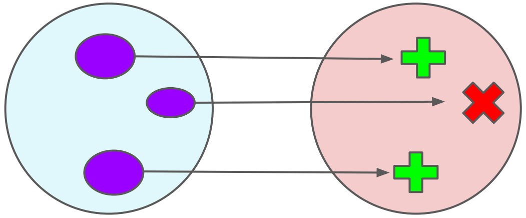
Preliminary Clustering

- Group points associated with same defect
- Identify related points of varying type



Label Generation

- Pair clusters with in-drive data points
- Non-trivial due to compounded measurement variance & calibration drift



Supervised Learning

- Determine if clusters of points indicate real defects