The reliable functioning of software depends crucially on tests (unit tests, input validation).

Despite their power, many machine learning models are sensitive to shifts in the data distribution.

In practice, ML pipelines rarely inspect incoming data for any signs of distributional shift.

Best practices for detecting shift in high-dimensional and real-life data have not yet been established.

Multiple univariate tests and multivariate kernel tests offer comparable detection performance.

BBSDs (univariate) and UAE (multivariate) are the best-performing shift detectors, respectively.

Top different samples from Classifiers are helpful in characterizing a shift’s nature and malignancy.

MNIST original split is not i.i.d., but harmless.

In this empirical study, we aim to identify the best dimensionality reduction technique and shift detector for detecting different shifts in MNIST and CIFAR-10.

Shift Detection Pipeline Overview

(a) Detection accuracy of different dimensionality reduction techniques across all simulated shifts on MNIST and CIFAR-10. Green bold entries indicate the best DR method at a given sample size, red italic shifts as harmful.

(b) Detection accuracy of different shifts on MNIST and CIFAR-10 using the best DR techniques. Green bold shifts are identified as harmless, red italic shifts as harmful.

Key Results

(b) Principal Components (PCA) are sensitive to shifts in the data distribution.

The reliable functioning of software depends crucially on tests (unit tests, input validation).

In practice, ML pipelines rarely inspect incoming data for any signs of distributional shift.

Best practices for detecting shift in high-dimensional and real-life data have not yet been established.

Multiple univariate tests and multivariate kernel tests offer comparable detection performance.

BBSDs (univariate) and UAE (multivariate) are the best-performing shift detectors, respectively.

Top different samples from Classifiers are helpful in characterizing a shift’s nature and malignancy.

MNIST original split is not i.i.d., but harmless.