UNIVERSITÀ DEGLI STUDI DI TRIESTE

# DISTRIBUTION SHIFT

## A Study on Their Effects on Statistical Models and Strategies for Mitigation

---


Andrea Spinelli, Giacomo Amerio,
Giovanni Lucarelli, Tommaso Piscitelli

University of Trieste

# Table of contents

# Introduction

# Dataset shift

- **Dataset shift** is a common problem in machine learning.
- It occurs when the distribution of the training data differs from the distribution of the test data.
- It is releted to another filed of study "**transfer learning** "
- This can lead to a decrease in the performance of the model.

# Most common cause of dataset shift

The two most common and well-studied causes of dataset shift are:

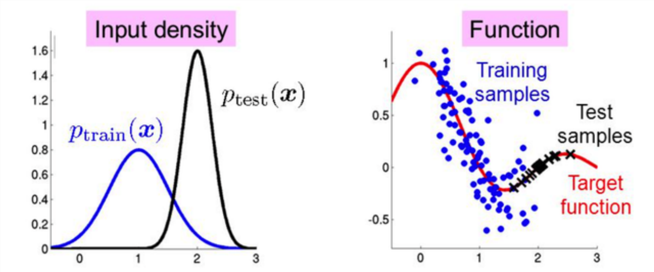
- Sample selection bias
- non stationary environments

In our project, we focused on one of the various forms of Dataset shift, **Covariate shift**, that is one of the most extensively reserched.

# Covariate shift

Can be formally defined as follows. Consider an input variable $X$ and a response variable $Y$, where $X \to Y$ represents the relationship between the two. Let $P_{\text{tra}}$ denote the probability distribution of the training data and $P_{\text{tst}}$ denote the probability distribution of the test data. A covariate shift occurs when:

$$P_{\text{tra}}(Y \mid X) = P_{\text{tst}}(Y \mid X) \quad \text{but} \quad P_{\text{tra}}(X) \neq P_{\text{tst}}(X)\,.$$

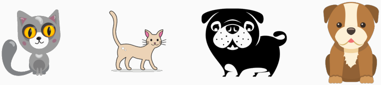


# Example

Consider a model designed to distinguish between cats and dogs:

**Training set:**


**Test set:**



- Model will not accurately distinguish between cats and dogs because the feature distribution will differ.
- Changes in the input distribution can significantly impact the model's accuracy.

# Inaccurate model

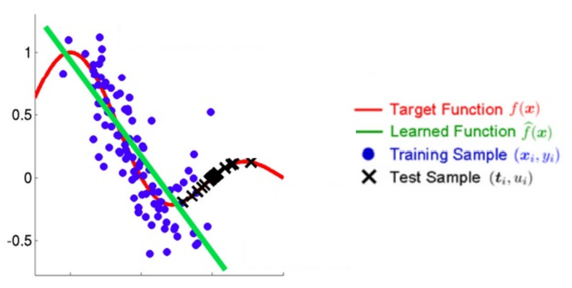



**Figure 1:** Example of inaccurate model.

In this study, we analyze the effects of distribution shift on different statistical models and propose strategies for its mitigation.

# Data Generation

# Training Dataset: Features

The dataset consists of $n = 10^4$ observations with 3 features and 1 target variable.

# Training Dataset: Features

The dataset consists of $n = 10^4$ observations with 3 features and 1 target variable.

**Features:**

- $X_{\text{train}} = (X_1, X_2, X_3) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\mu_i \sim \mathcal{U}_{[0,1]}$ for $i = 1, 2, 3$
- $[\boldsymbol{\Sigma}]_{i,j} \sim \mathcal{U}_{[-1,1]}$ for $i, j = 1, 2, 3$

**Note:** The $\boldsymbol{\Sigma}$ randomly generated has been transformed to a symmetric and positive semidefinite matrix by computing $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T$.

## Training Dataset: Target Variable

Dicothomous target variable: $Y \in \{0, 1\}$

$$z = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i=1}^{3} \beta_{ii} x_i^2 + \sum_{i=1}^{2} \sum_{j=i+1}^{3} \beta_{ij} x_i x_j$$

$$Y \sim \mathrm{Be}(p)\,, \quad p = \frac{1}{1 + e^{-z}}$$

# Testing Dataset

Same dataset structure as the train set, but:

- $X_s = (X_{1s}, X_{2s}, X_{3s}) \sim \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$
- $\boldsymbol{\mu}_s = \mathcal{Q}_{0.05}(X_{\text{train}})$
- $\boldsymbol{\Sigma}_s$ generated in the same way as $\boldsymbol{\Sigma}$ (but $\boldsymbol{\Sigma}_s \neq \boldsymbol{\Sigma}$)

# Performance Degradation

# Performance Enhancement

Questions?

# References i