

What is covariate shift?

MONITORING MACHINE LEARNING CONCEPTS



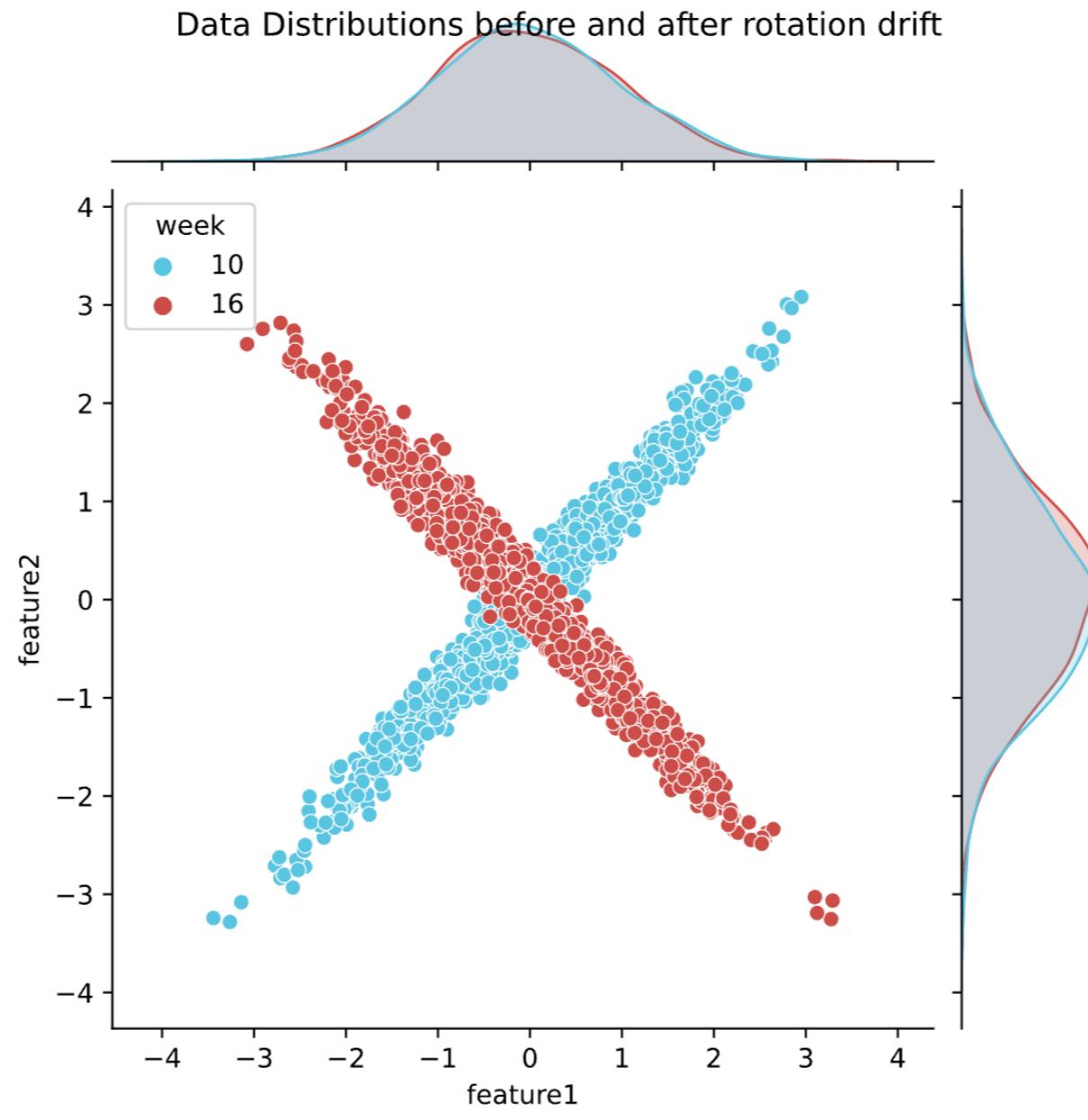
Hakim Elakhrass

Co-founder and CEO of NannyML

Definitions

- covariate variables = input features
- $P(X)$ changes
- joint probability $P(Y|X)$ remains the same
- changes in the joint distribution of the covariates

Why joint probability distribution?



Why does covariate shift occur?

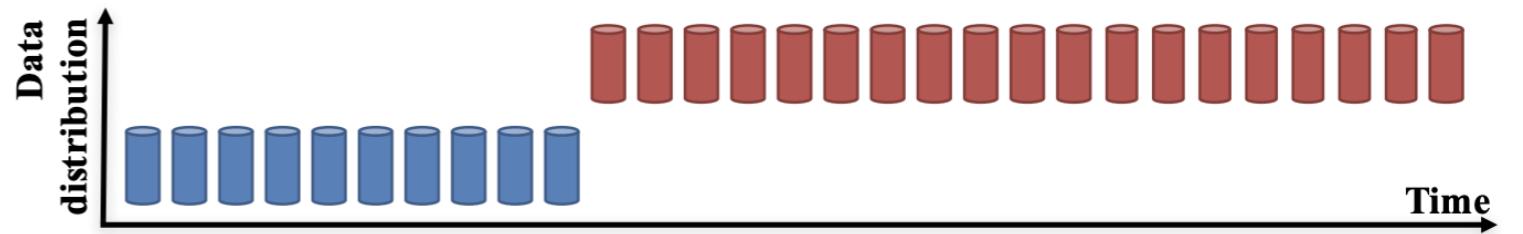
Potential reasons for covariate shift:

- The real world is not stationary - patterns and trends evolve
- Changes in data sources - variations in how data is collected between testing and production
- Evolution of the system and environment

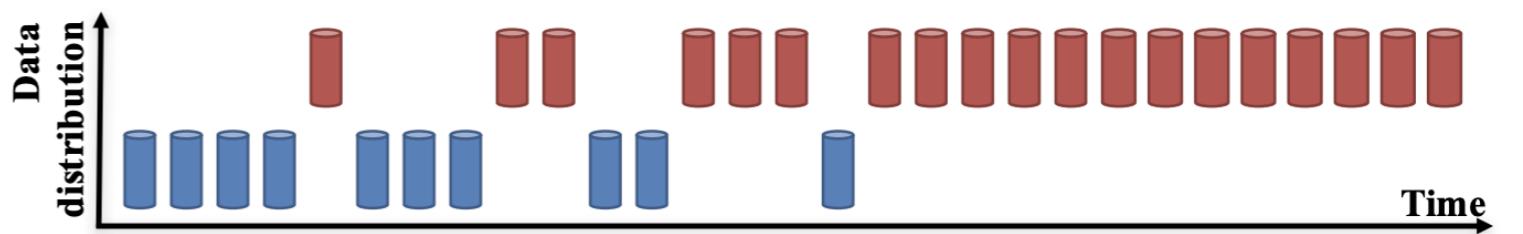
How does covariate shift occur?

Dynamics of the changes in the distribution:

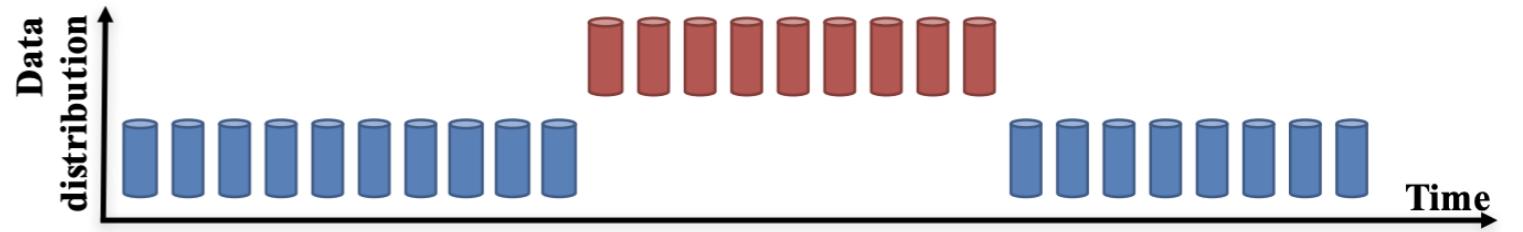
- Sudden



- Gradual

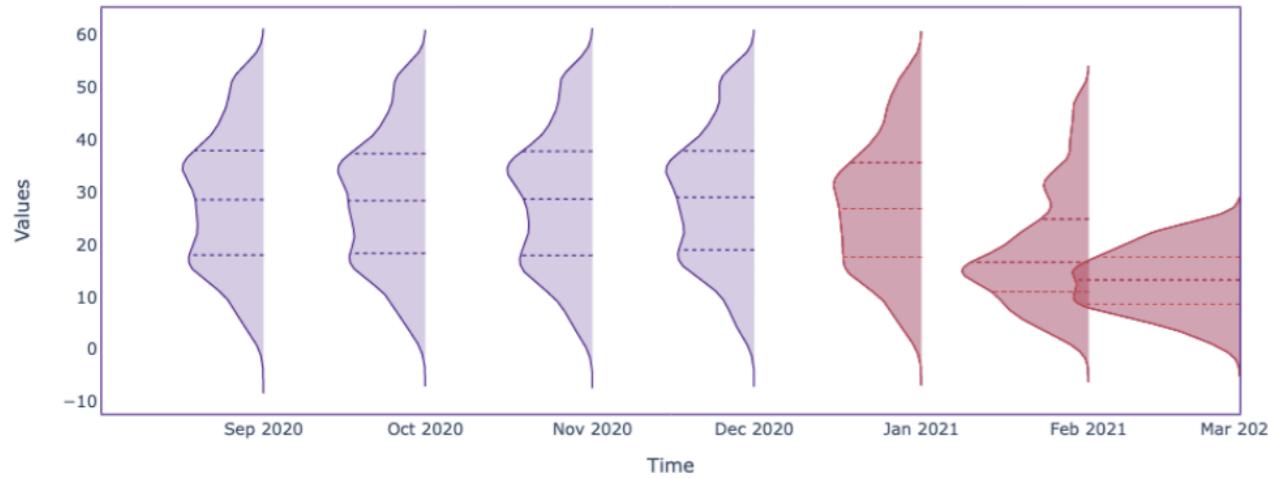


- Seasonal

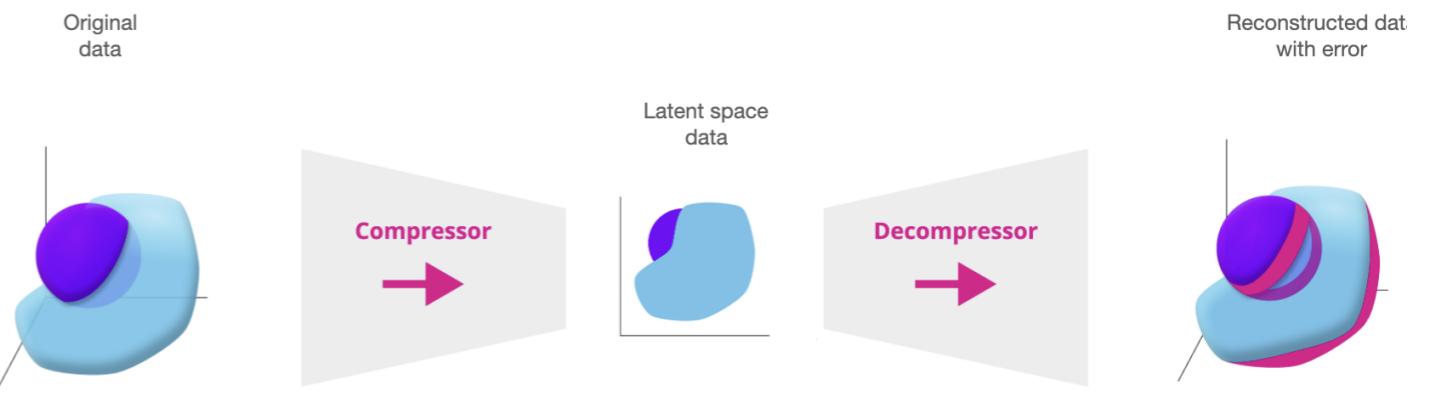


How to detect the covariate shift?

Univariate method



Multivariate method



¹ <https://app.datacamp.com/learn/courses/dimensionality-reduction-in-python>

Let's practice!

MONITORING MACHINE LEARNING CONCEPTS

How to detect covariate shift

MONITORING MACHINE LEARNING CONCEPTS

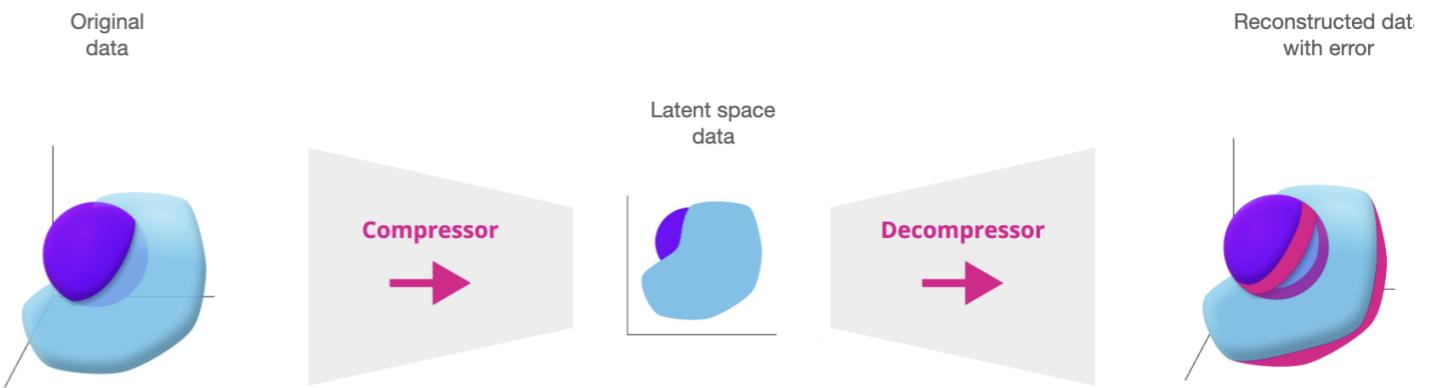


Hakim Elakhrass

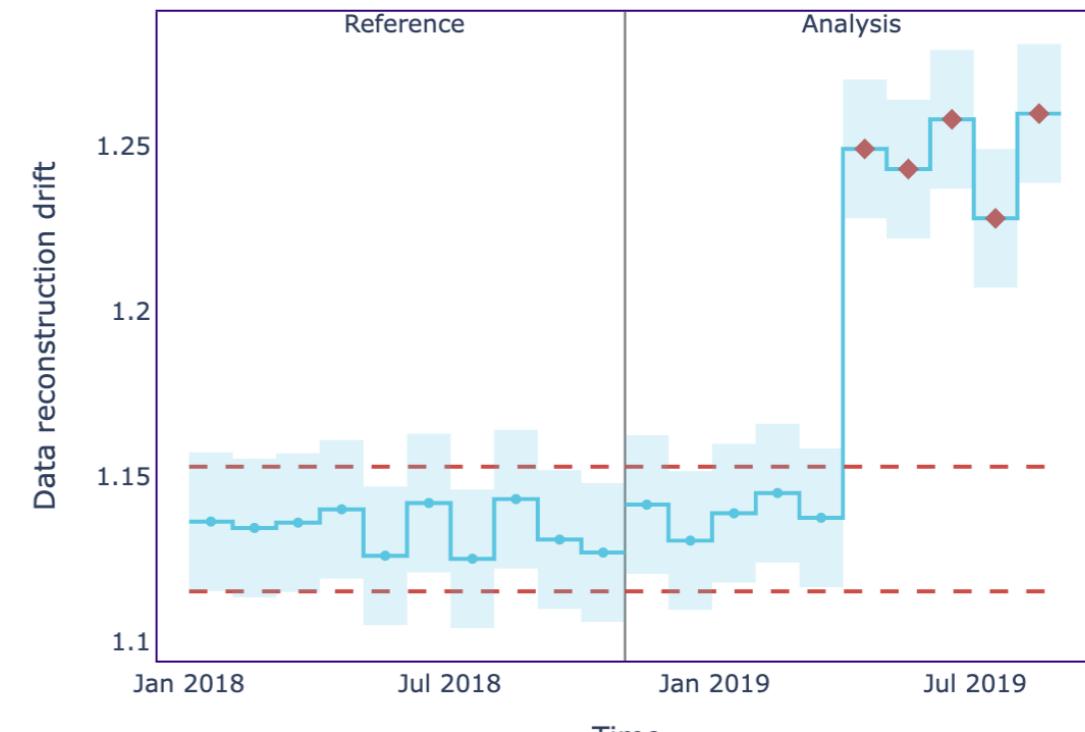
Co-founder and CEO of NannyML

Multivariate drift detection

- Looks for changes in joint distribution



- Uses the PCA algorithm for data compression
- Uses reconstruction error as a measure of drift



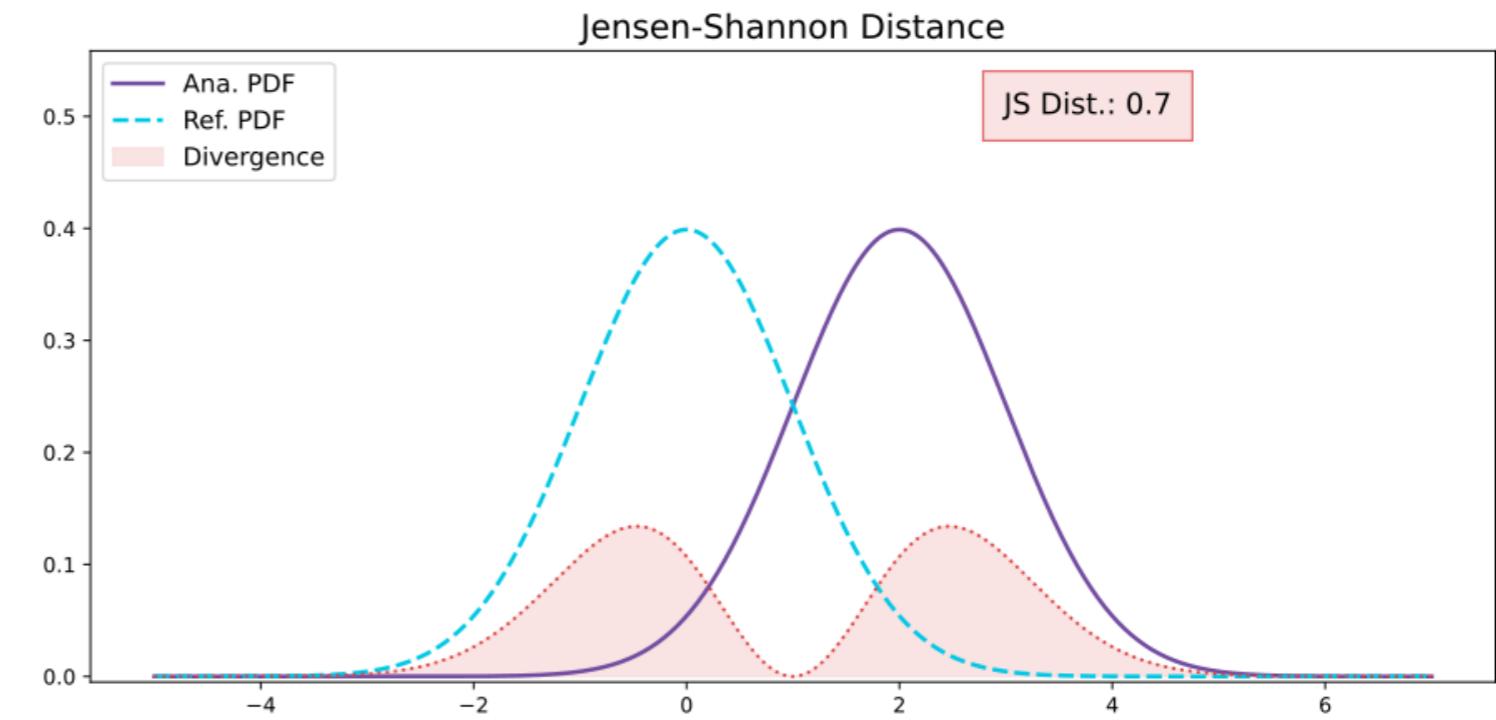
Univariate drift detection

Types of variables:

- Categorical - represent types of data which may be divided into groups like martial status, smoking status, level of education
- Continuous - a variable with an infinite number of real values within a given interval like height, weight, distance, time

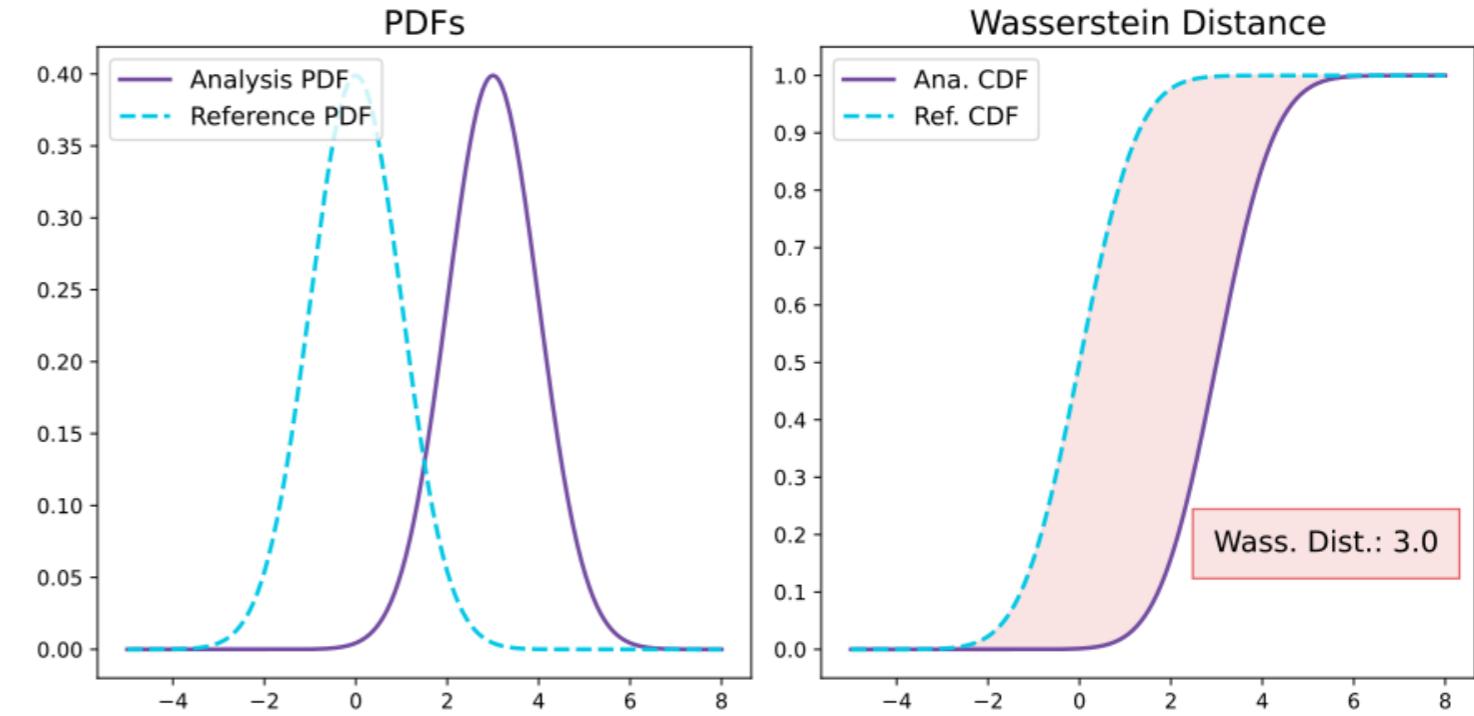
Continuous methods - Jensen-Shannon

- Measures the similarity of two distributions
- Range $[0, 1]$
- Catches meaningful low-magnitude drifts



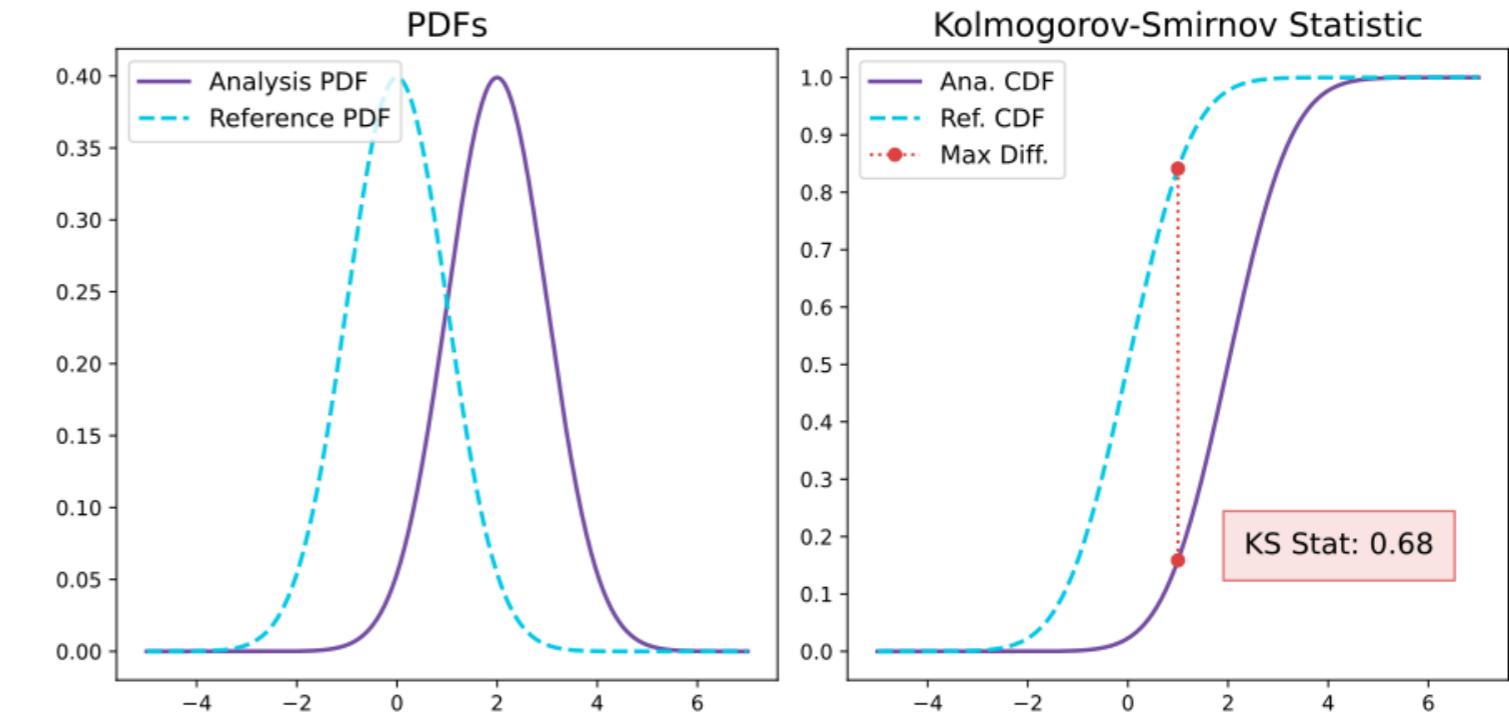
Continuous methods - Wasserstein

- The minimum effort needed to transform one distribution into another
- Range $[0, +\infty]$
- Sensitive to outliers



Continuous methods - Kolmogorov-Smirnov

- Maximum distance of the cumulative distribution functions
- Range $[0, 1]$
- Prone to false positives

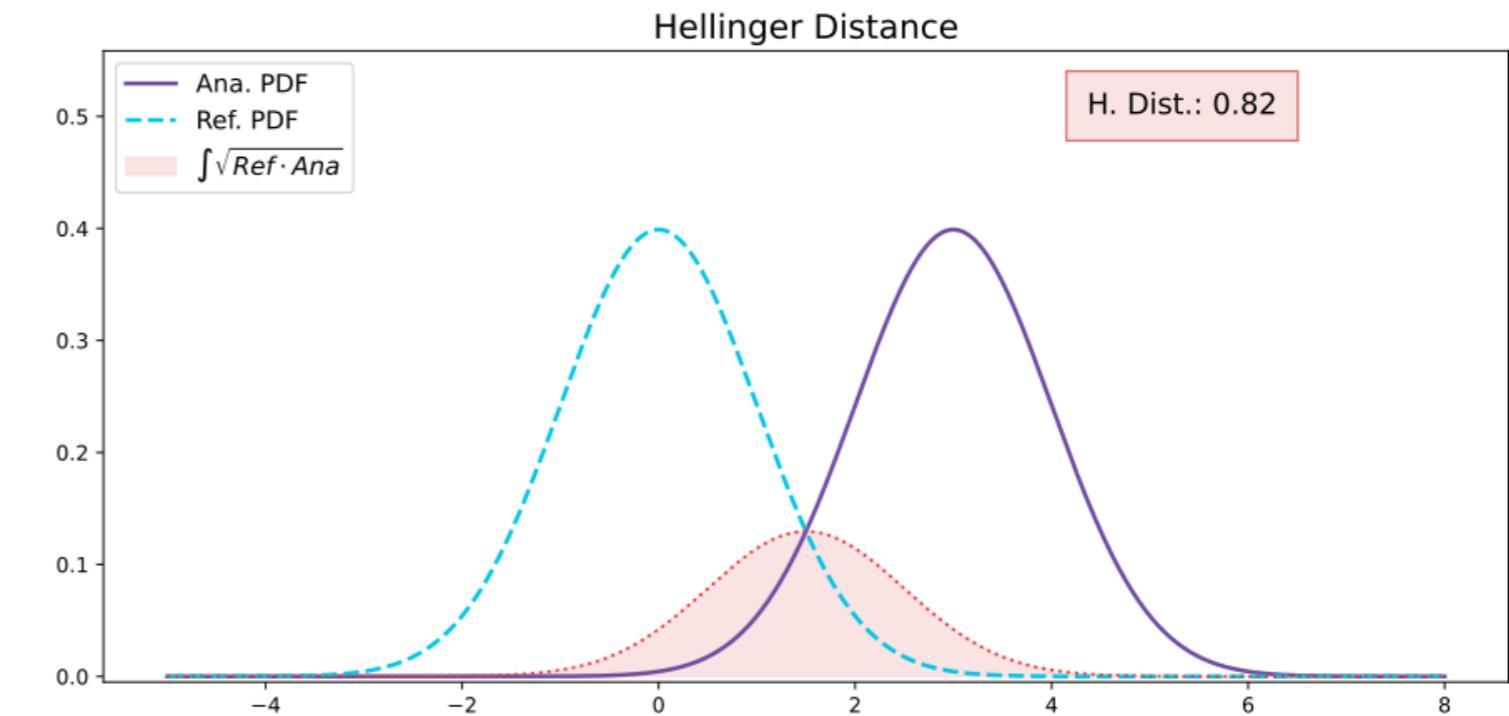


Continuous methods - Hellinger

- Overlap between distributions
- Range $[0, 1]$
- Doesn't differentiate between strong shifts

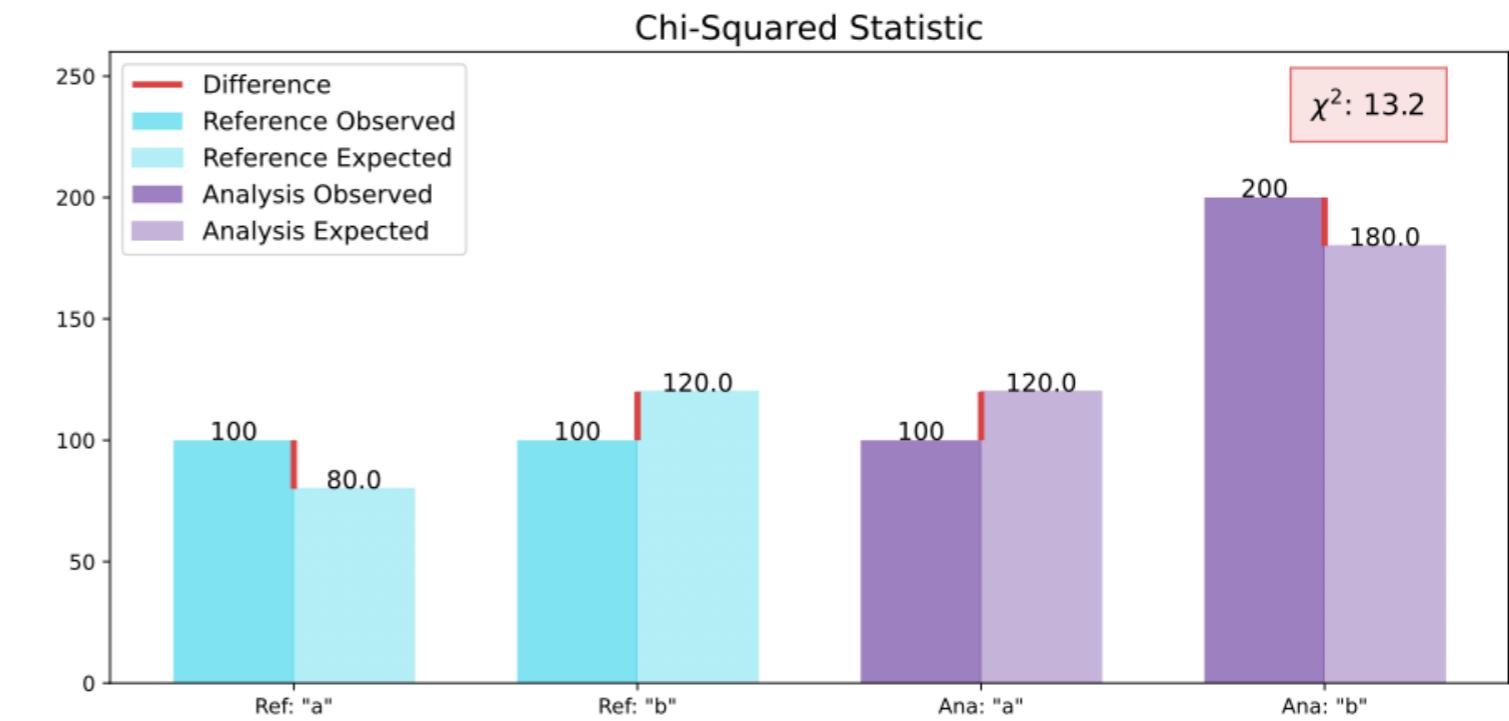
Continuous methods - Recommendation

- Jensen-Shannon and Wasserstein generally perform well



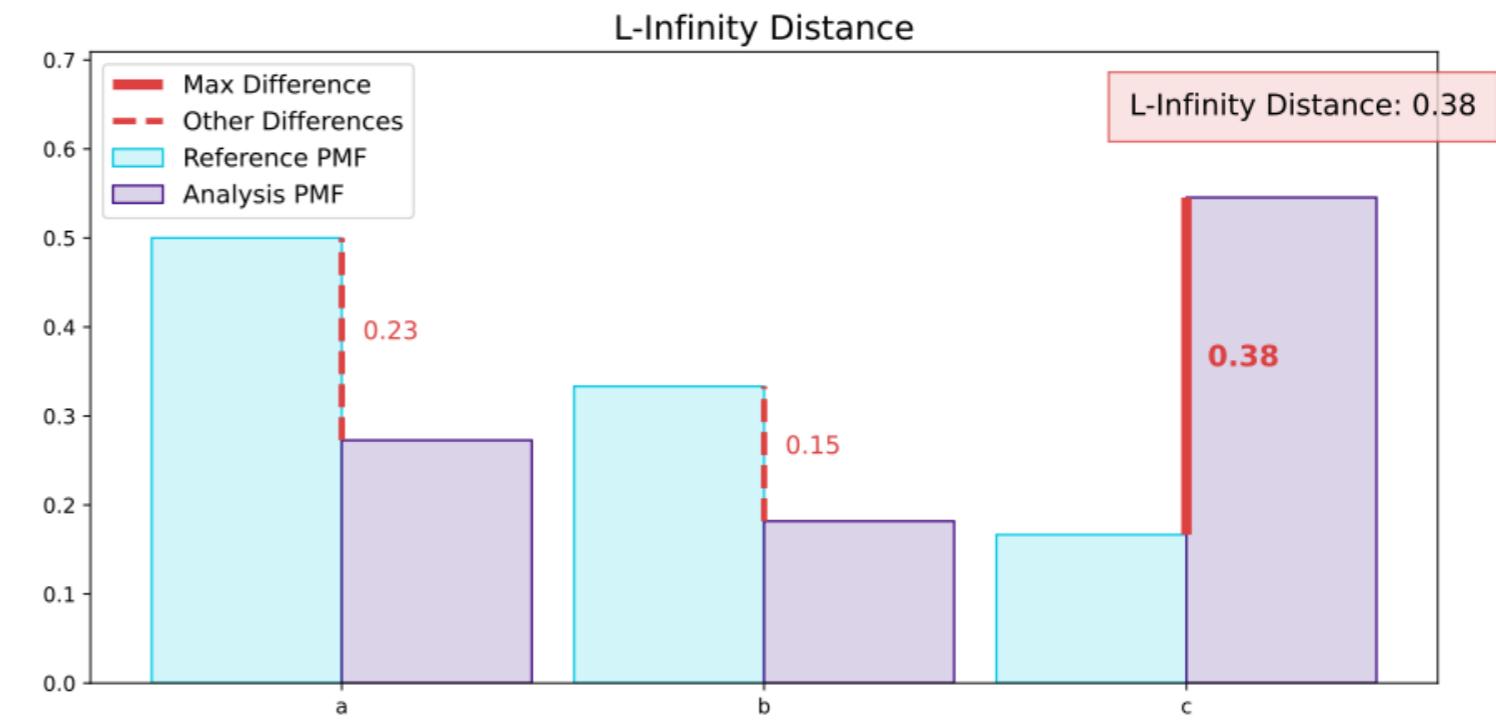
Categorical methods - Chi-squared

- Sensitive in changes for low-frequency categories



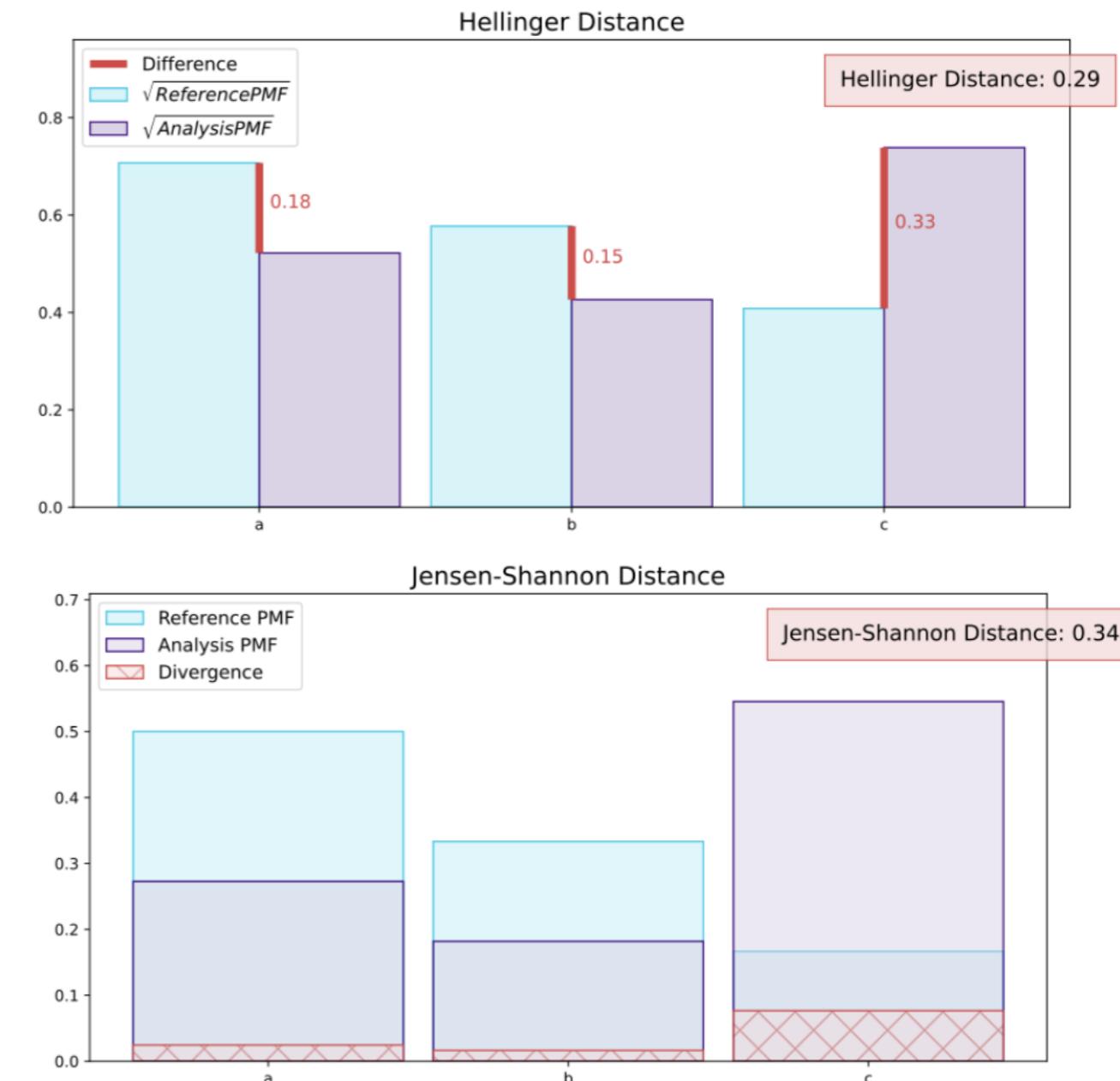
Categorical methods - L-infinity

- Identifies the most significant shift across all categories



Categorical methods - Jensen-Shannon and Hellinger

- Jensen-Shannon or L-Infinity when dealing with many categories
- L-Infinity distance to detect changes in individual categories



Let's practice!

MONITORING MACHINE LEARNING CONCEPTS

What is concept drift?

MONITORING MACHINE LEARNING CONCEPTS

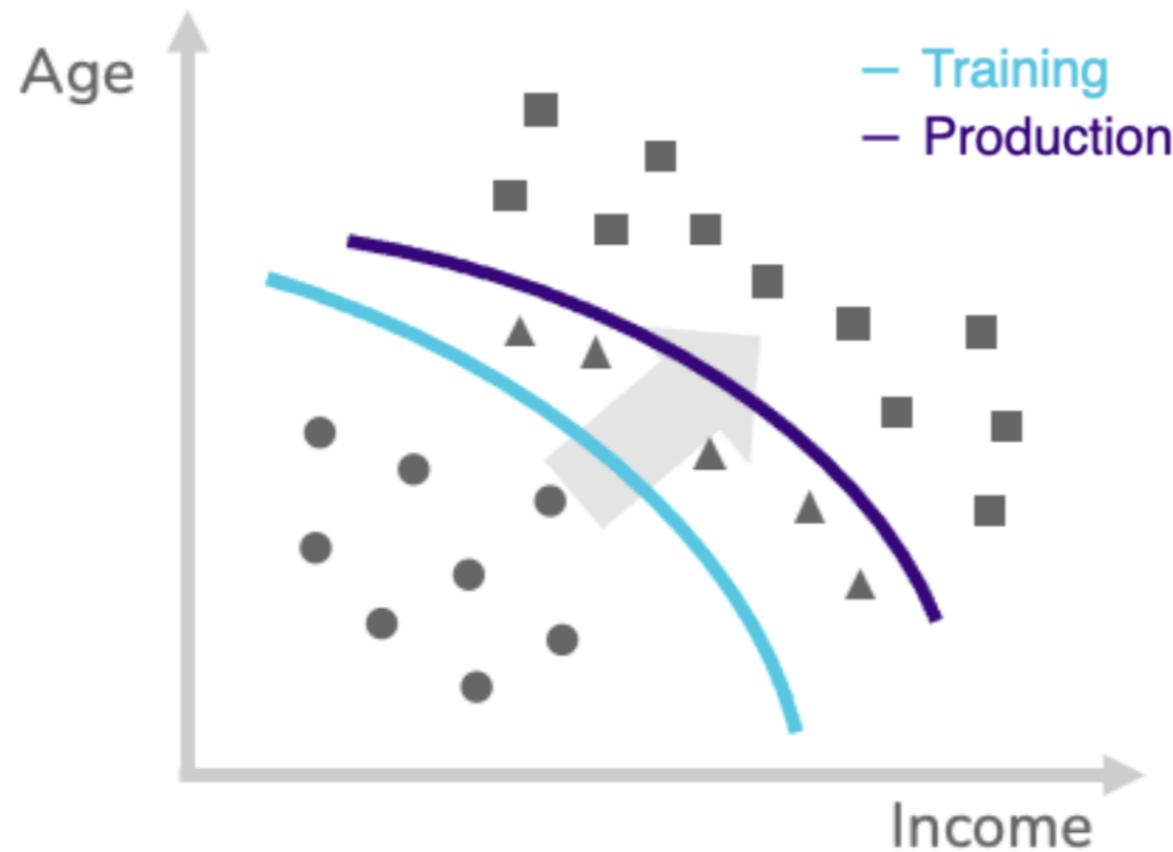


Hakim Elakhrass

Co-founder and CEO of NannyML

Definition

- Change in relationship between the model inputs and the target
- $P(Y|X)$ changes, $P(X)$ stays the same

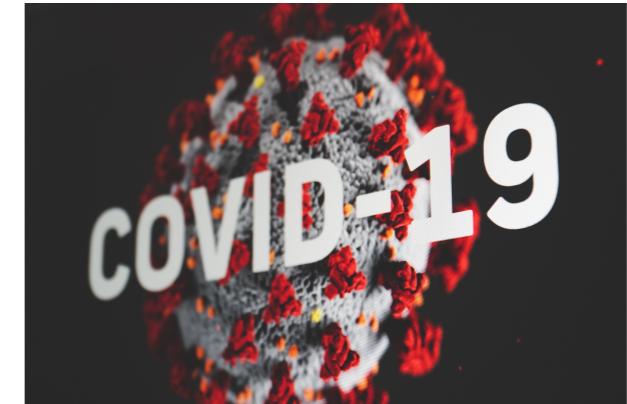


Why drift happens?

- External events
 - viral trends, policy changes
- Unmodeled seasonality
 - in case of demand forecasting seasonal events like Black Friday or Christmas
- Changes in data-generation process
 - new update to the data collection app
- Evolving user behavior
 - habits, patterns, preferences are constantly changing

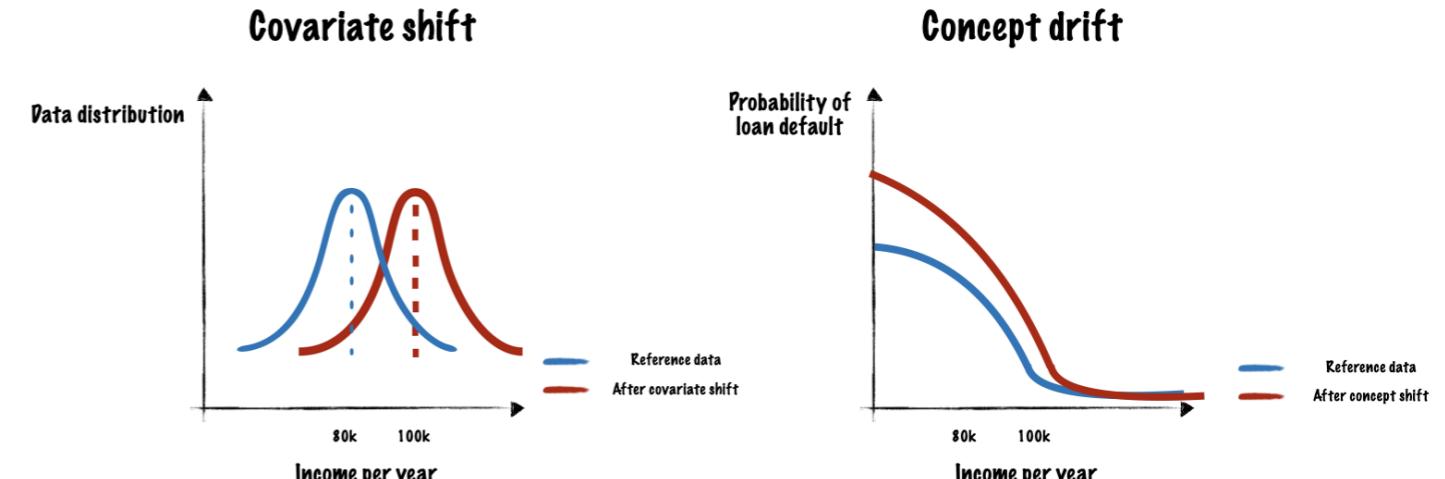
The dynamics of concept drift

- Sudden drift - a new concept occurs within the short time
- Gradual drift - a new concept gradually replaces the old one
- Reoccurring - reoccurring old concept over time

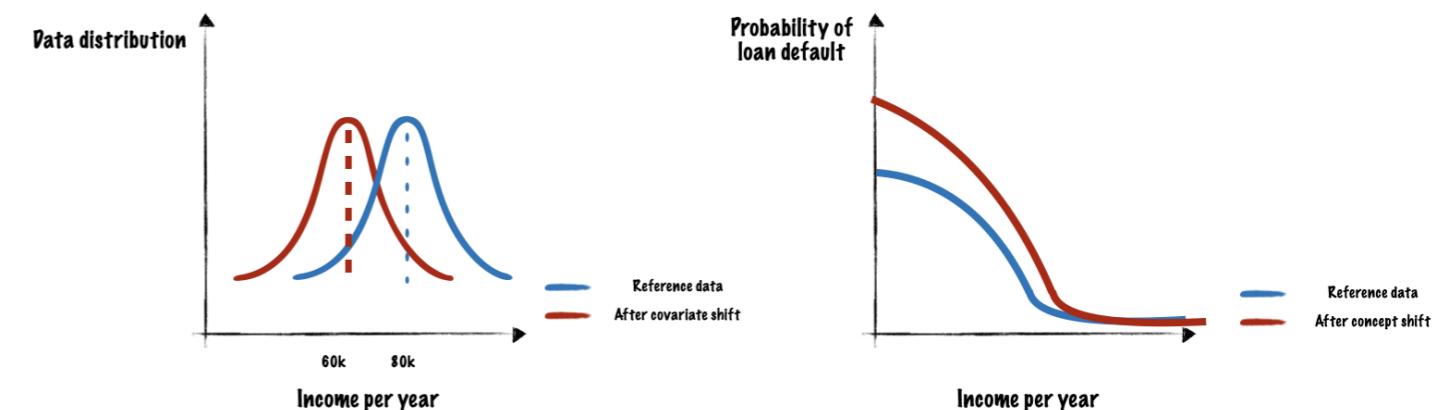


Effects of covariate shift on concept drift

- Negative
 - the effect of concept drift decreases



- Positive
 - the effect of concept drift intensifies



Let's practice!

MONITORING MACHINE LEARNING CONCEPTS

How to handle concept drift?

MONITORING MACHINE LEARNING CONCEPTS



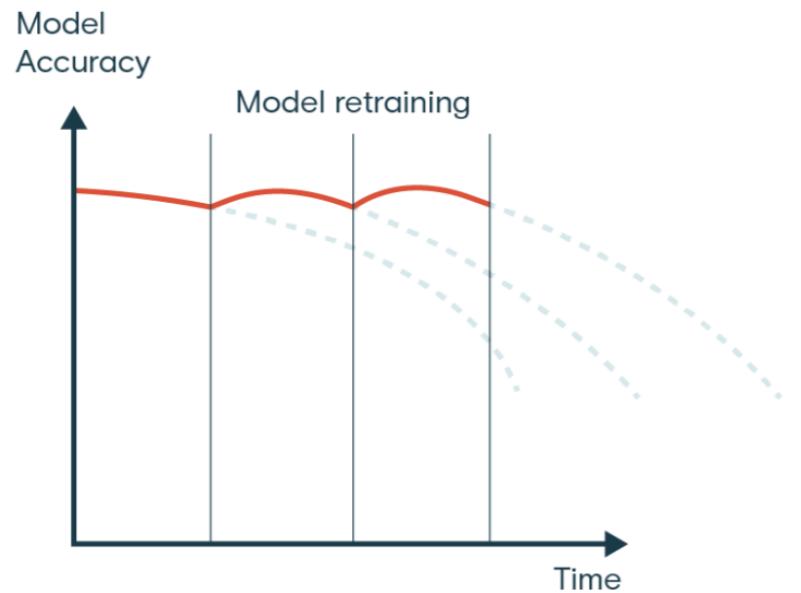
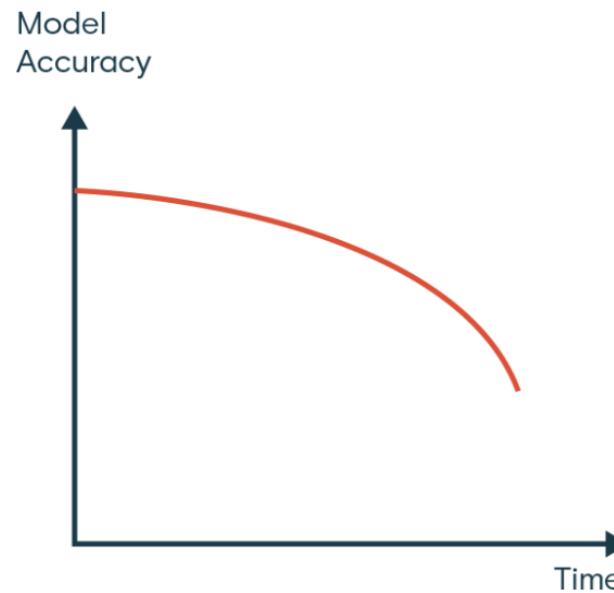
Hakim Elakhrass

Co-founder and CEO of NannyML

Concept drift detection

- Error-based methods
 - tracking error changes over time
 - requires ground truth
- Train a new model using training and production data
 - change in the predictions is a concept drift
 - expensive in more advanced use-cases

Retraining



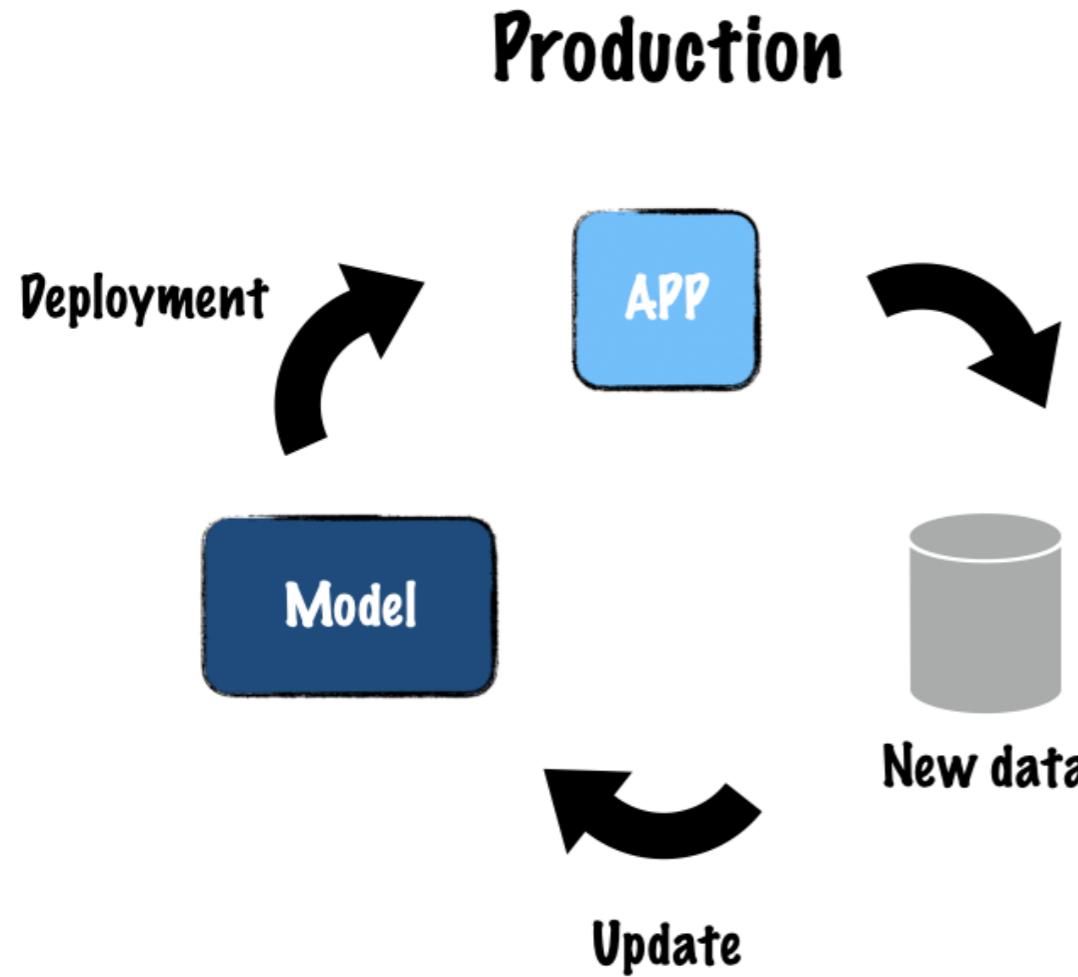
Pros :

- keep the model up-to-date with recent patterns

Cons :

- increased costs and risk of failure
- doesn't provide the root cause of the problem

Online learning



Pros :

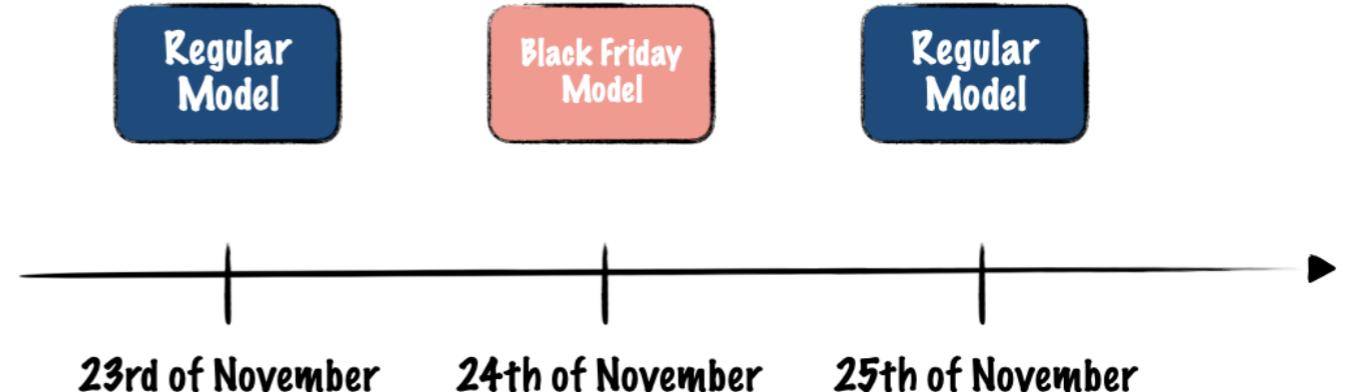
- real-time adaptation to changing conditions

Cons :

- requires constant access to ground truth
- sensitive to noise
- needs careful parameter tuning

Other resolutions

- A event-specific model for reoccurring events
- Weighting the importance of new data
 - with most focus on newer data, model can adapt easier



Let's practice!

MONITORING MACHINE LEARNING CONCEPTS

Wrap-up

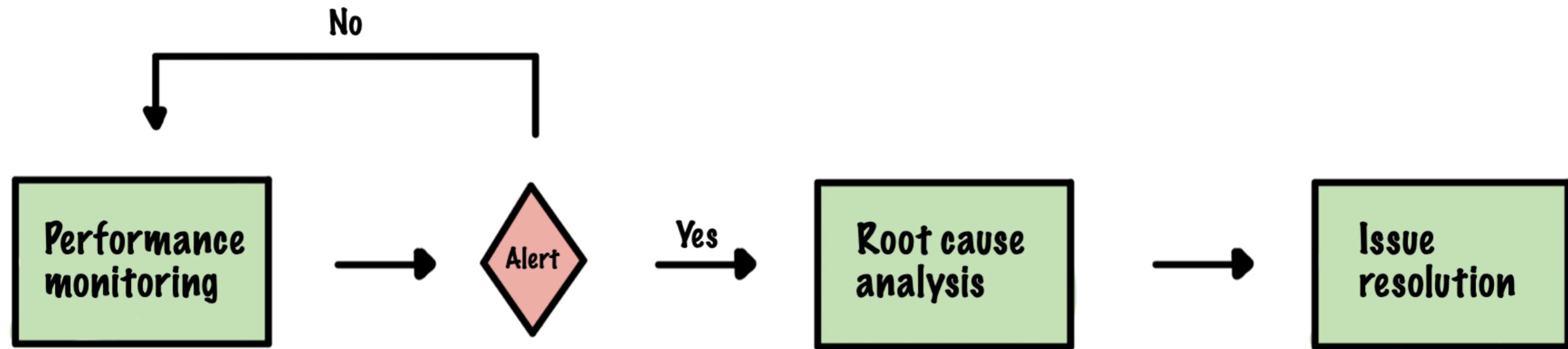
MONITORING MACHINE LEARNING CONCEPTS



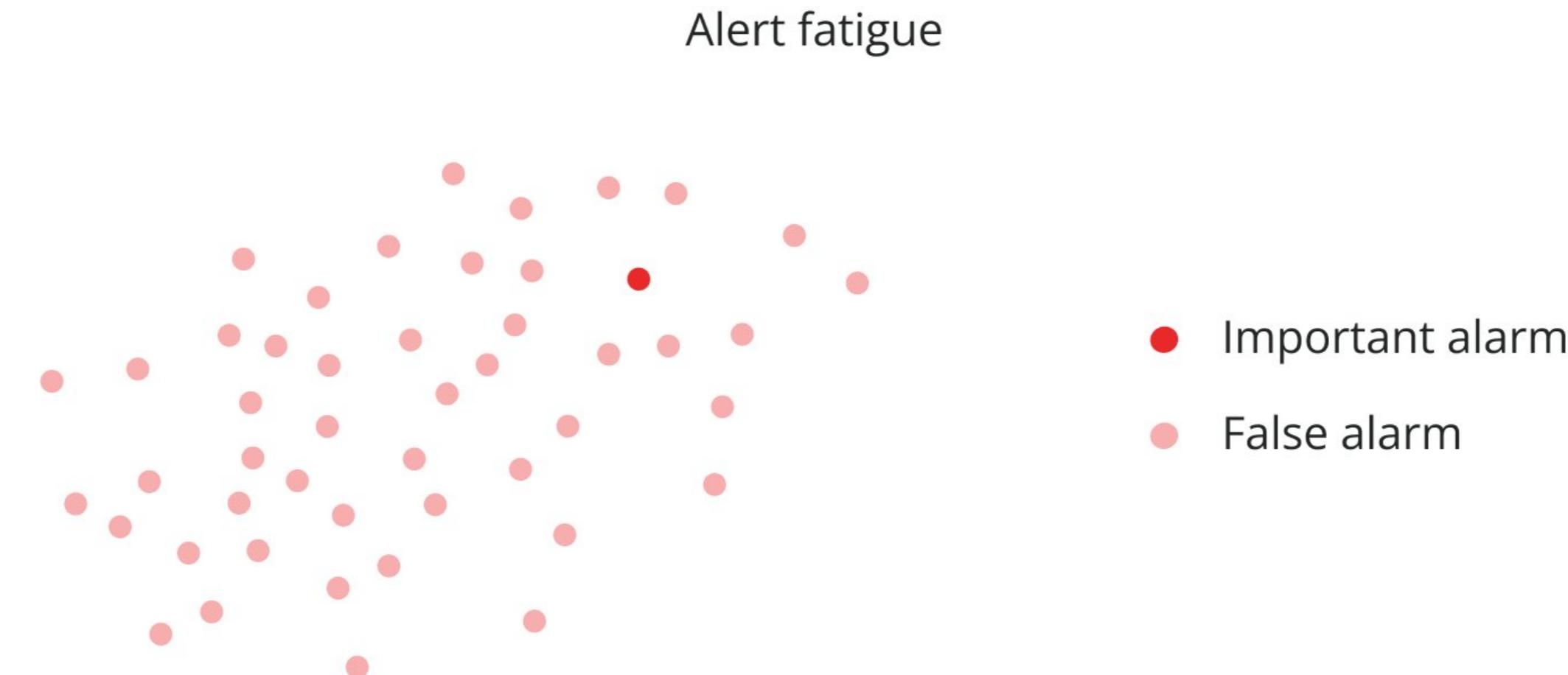
Hakim Elakhrass

Co-founder and CEO of NannyML

Chapter 1 - What Is ML Monitoring?

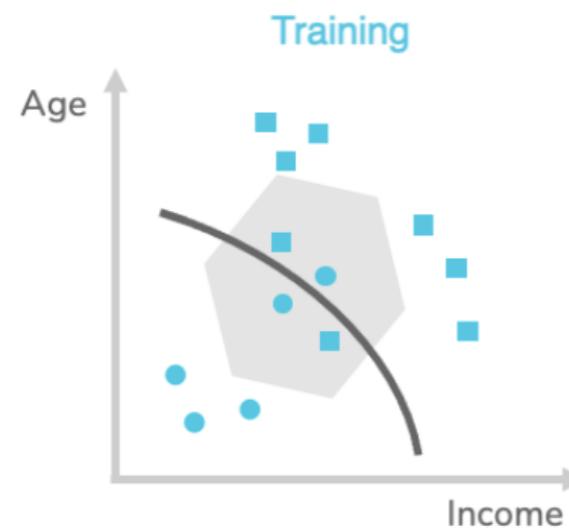


Chapter 2 - Theoretical Concepts of Monitoring

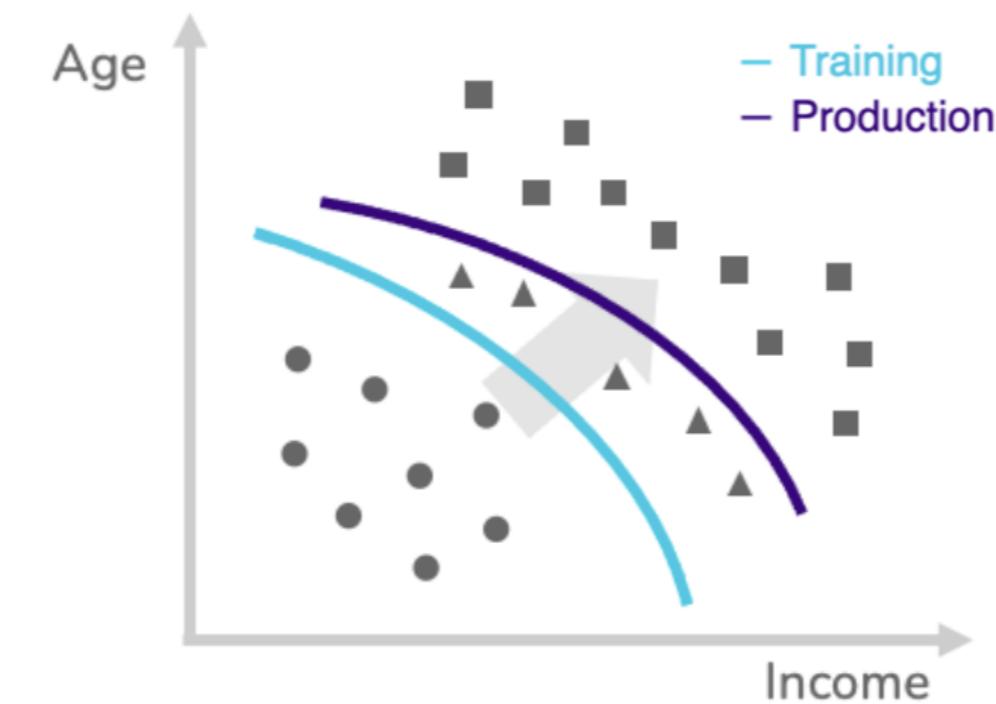


Chapter 3 - Covariate Shift and Concept Drift

Covariate shift



Concept drift



Congratulations!

MONITORING MACHINE LEARNING CONCEPTS