ILAB: OeNB

Documentation

Variational Autoencoders

Author:
Tristan Leiter

13. Februar 2026

Inhaltsverzeichnis

Kapitel 1

Hybrid VAE-XGBoost Architecture & Implementation

1.1 Motivation and Objective

The primary objective of this study is to augment a standard Gradient Boosting Machine (XGBoost) with unsupervised representation learning. While XGBoost is a powerful supervised learner, it relies on orthogonal decision boundaries that may fail to capture the complex, non-linear manifolds typical of distressed firms (e.g., "Zombie companies" that are technically insolvent but survive via liquidity).

We adopt a "Hybrid Expert" architecture where a Variational Autoencoder (VAE) and a Denoising Autoencoder (DAE) act as upstream feature extractors. These neural networks compress high-dimensional financial data into dense representations, which are then fed into the XGBoost model alongside the original features.

1.2 Implementation Strategies

The implementation iterates through four distinct feature engineering strategies (A through D), each targeting a specific type of risk signal.

1.2.1 Strategy A: Latent Manifold Features (Standard VAE)

This strategy utilizes the **Encoder** of a standard Variational Autoencoder to perform non-linear dimensionality reduction.

- **Implementation:** The model compresses the 11 continuous financial ratios and one-hot encoded categorical metadata into an 8-dimensional probabilistic latent space.
- **Mechanism:** We extract the mean vector μ of the latent distribution $P(z|x)$ for each observation.
- **Code Logic:**
  - The VAE is trained to minimize the Evidence Lower Bound (ELBO), balancing reconstruction loss against the KL-divergence from a standard normal prior.
  - `Strategy_A_LF` = The 8-column matrix output from the bottleneck layer ($l_1 \dots l_8$).
- **Hypothesis:** These features capture "clusters of behavior" (e.g., high-growth/low-liquidity) that raw ratios cannot express individually.

1.2.2 Strategy B: Anomaly Scores (Reconstruction Error)

This strategy utilizes the **Decoder** to quantify how "weird" a firm's financial structure is relative to the population.

- **Implementation:** We measure the distance between the original input vector x and the VAE's reconstruction $\hat{x}$. To handle mixed data types (continuous and categorical), the code implements a split-loss calculation.

- **Formula (Balanced Score):**

$$\text{Score} = w_{cont} \cdot \left(\frac{1}{N_{cont}} \sum (x_{cont} - \hat{x}_{cont})^2 \right) + w_{cat} \cdot \left(\frac{1}{N_{cat}} \text{BCE}(x_{cat}, \hat{x}_{cat}) \right) \tag{1.1}$$

  Where BCE is the Binary Cross Entropy for categorical variables.

- **Code Logic:**
  - `anomaly_score_cont_avg`: MSE normalized by the number of continuous columns.
  - `anomaly_score_cat_avg`: BCE normalized by the number of categorical columns (one-hot levels).
  - The normalization prevents the score from being dominated by the categorical vector simply due to its higher dimensionality.
- **Hypothesis:** High reconstruction error implies the firm's financial structure deviates from the norm. In credit risk, structural outliers are often highly correlated with default (e.g., fraud or extreme distress).

1.2.3 Strategy C: Robust Features (Denoising Autoencoder)

This strategy moves beyond the standard VAE by implementing a **Denoising Autoencoder (DAE)** to force feature robustness.

- **Implementation:** Unlike Strategy A, this model is trained by intentionally corrupting the input data while forcing the model to predict the clean original data.
- **Mechanism:**
  - **Noise Injection:** A Gaussian Noise layer ($\sigma = 0.1$) is added immediately after the input layer in Keras.
  - **Training Objective:** Minimize $L(x, g(f(\tilde{x})))$, where $\tilde{x} = x + \epsilon$ and $\epsilon \sim \mathcal{N}(0, 0.1)$.
- **Code Logic:**
  - A separate Keras model (`dae_autoencoder`) is defined using the Functional API.
  - The encoder weights are extracted after training to generate features `dae_l1` ... `dae_l8`.
- **Hypothesis:** By forcing the network to "subtract" the noise, the latent features become invariant to small fluctuations and measurement errors in the financial ratios, focusing only on the robust structural signal.

1.2.4 Strategy D: Domain Expert Interactions (The "Zombie" Detectors)

This strategy eschews deep learning in favor of explicit domain knowledge, specifically targeting "Zombie Companies"—firms that are profitable but insolvent, or solvent but illiquid.

- **Implementation:** We manually construct interaction terms that define diagonal decision boundaries XGBoost might struggle to approximate with shallow trees.
- **Key Features Created:**
  - **Distress Gap (`Gap_Debt_Equity`):** $f_{11} - f_6$. This measures the absolute distance between Liabilities and Equity. A high positive gap indicates a leverage crisis regardless of firm size.
  - **Cash Burn Ratio (`Ratio_Cash_Profit`):** $f_5/(|f_8| + \epsilon)$. This identifies "Profitable but Illiquid" firms (False Negatives) where high accounting profit masks a dangerously low cash position.
  - **Feature Stabilizer:** A conditional feature (`ifelse`) that swaps Profit for Cash when Profit is negative, creating a continuous "Solvency Capacity" metric.
- **Hypothesis:** These ratios explicitly expose the "hidden buffers" that allow distressed firms to survive, directly addressing the "Healthy Loser" confusion matrix quadrant.

1.3 Base Model Integration

The final modeling stage aggregates the outputs of these strategies:

$$X_{final} = [X_{raw}, X_{Latent(A)}, X_{Anomaly(B)}, X_{Robust(C)}, X_{Expert(D)}] \tag{1.2}$$

This augmented dataset is fed into an XGBoost classifier. The inclusion of VAE/DAE features allows the gradient boosting model to view the data through multiple "lenses": the raw financial values, the probabilistic manifold (VAE), the robust structural view (DAE), and the financial analyst's view (Strategy D).

Kapitel 2

Final Model Optimization: The Solvency Stabilizer

2.1 The Breakthrough: Solving the "Healthy Loser" Panic

The primary weakness identified in previous iterations was the model's tendency to panic when observing negative profitability, resulting in a high rate of False Positives. This phenomenon, termed the "Healthy Loser" Panic, flagged firms that were technically unprofitable but structurally sound due to high cash reserves.

By implementing **Strategy D**, which utilizes the `Feature_Stabilizer` (switching the focus to Cash $f5$ when Profit $f8$ is negative), we achieved a "Double Win": the model successfully caught more defaulters while drastically reducing false alarms.

2.2 Quantitative Impact: Before vs. After

The implementation of the stabilizer logic produced a decisive shift in model performance. Table 2.1 details the impact on the confusion matrix.

Tabelle 2.1: Impact of Strategy D (Feature Stabilizer) on Model Performance

Metric	Previous Result	Strategy D Result	Improvement
True Positives (Caught Risks)	1,191	**1,361**	+170 (Sensitivity ↑)
False Negatives (Missed Risks)	275	**105**	-170 (Safety ↑)
False Positives (False Alarms)	39,738	**26,077**	-13,661 (Precision ↑)
True Negatives (Correct Safe)	129,450	**143,111**	+13,661 (Efficiency ↑)

Key Outcomes:

- **Efficiency:** 13,661 healthy firms were rescued from being wrongly flagged as risky.
- **Sensitivity:** Simultaneously, an additional 170 defaulters were caught that were previously missed.
- **Total Capture Rate:** The model now captures **92.8%** of all defaults ($1,361/1,466$), a figure considered exceptional for credit risk modeling.

2.3 Mechanism of Action: Why the Stabilizer Worked

The significant drop in False Positives (from 39k to 26k) confirms the hypothesis regarding "Healthy Losers."

- **The Old Logic:** The model observed $f8$ (Profit) at values like -2.0 and immediately classified the firm as high risk based on income statement bleeding.
- **The New Logic:** The `Feature_Stabilizer` intervened: *"Profit is negative; therefore, evaluate Cash ($f5$) instead."*

- **The Result:** For the 13,661 rescued firms, the model recognized the "Cash Cushion" ($f5 > 0$) as a sufficient buffer against the "Profit Bleed," correctly reclassifying them as Survivors (True Negatives).

2.4 Forensic Analysis of Remaining Errors

Despite optimization, specific error groups remain. A forensic analysis reveals the distinct profiles of these residual errors.

2.4.1 The Remaining False Negatives (105 Firms)

These are the **"Stealth Defaulters."**

- **Profile:** Median Profit ($f8$) ≈ -0.157; Median Solvency Gap ≈ -0.235.
- **Diagnosis:** These firms exhibit "safe" balance sheets (Equity > Debt) and are barely losing money.
- **Root Cause:** Financially, they appear identical to safe firms. Their default is likely driven by non-financial factors—such as fraud, lawsuits, or sudden management exits—or extreme short-term liquidity shocks not captured in annual reports. This represents the likely "irreducible error" floor for a model based solely on annual financial statements.

2.4.2 The Remaining False Positives (26,077 Firms)

These are the **"Hardcore Zombies."**

- **Profile:** Massive Losses (Median $f8 \approx -1.05$); Massive Debt (Median Gap $\approx +0.90$).
- **Diagnosis:** By all standard financial logic, these firms should be insolvent.
- **Root Cause:** Their survival is likely due to external factors such as government bailouts, parent company guarantees, or extreme asset liquidation.
- **Interpretation:** Flagging these firms is the *correct behavior* for a risk model. They represent high risk; they simply survived against the odds.

2.5 Future Improvements and Limitations

While the current model is production-ready, further reduction of the remaining error types requires data beyond the current scope.

2.5.1 Addressing False Positives (The "Zombie" Survivor)

The model fails to exclude these firms because their survival is mathematically improbable based on their financials alone.

- **Limitation:** The model cannot see "external support."
- **Solution path:** To reduce these False Positives, we would need to integrate:
  1. **Ownership Structure Data:** Identifying parent companies with deep pockets.
  2. **State Aid/Subsidy Data:** Identifying firms receiving government support.
  3. **News Sentiment:** Detecting announcements of restructuring or bailouts.
- **Recommendation:** Treat these 26,077 firms as a "High Risk Watchlist." They are not defaults yet, but they are living on borrowed time.

2.5.2 Addressing False Negatives (The "Stealth" Defaulter)

The model fails to catch these firms because their annual reports look healthy right up until the moment of collapse.

- **Limitation:** The latency of annual reporting masks sudden liquidity crises or fraud.
- **Solution path:** To capture these Stealth Defaulters, we would need:
    1. **Higher Frequency Data:** Monthly cash flow or bank transaction data to catch sudden liquidity drying.
    2. **Fraud Detection Models:** Applying Benford's Law or forensic accounting ratios to detect manipulated financial statements.
    3. **Legal Filings:** Monitoring court dockets for sudden litigation which often precedes "healthy" defaults.

2.6 Final Recommendation

We have reached the point of diminishing returns for feature engineering on this specific dataset. The recommendation is to **Stop Engineering** and accept the model. Capturing ∼93% of defaults with a drastically reduced False Positive rate is a robust result. The pipeline (Strategy D + Threshold Optimization) should now be applied to the final Holdout Test Set to confirm these metrics on unseen data.