Momentum-Based Optimizers for Predictive Maintenance of A Turbofan Jet Engine

Peter Gauchat, Yumeng Xiu

April 16, 2025

1 Introduction

Equipment failure disrupts operations in industries ranging from design to manufacturing. Predicting the Remaining Useful Life (RUL) of industrial equipment is a critical task in predictive maintenance that can significantly reduce operational costs and prevent catastrophic failures. This report presents the implementation and evaluation of two machine learning approaches for RUL prediction: Support Vector Regression (SVR) and a Multi-Layer Perceptron (MLP) neural network with AdamW optimization. The models were trained and evaluated on the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset, specifically the FD001 subset, which contains run-to-failure data for aircraft engines under a single operating condition with a single failure mode.

2 Methodology

2.1 Dataset and Preprocessing

We utilized the NASA C-MAPSS dataset (FD001), which contains run-to-failure data for aircraft engines under a single operational condition and failure mode. The preprocessing included:

- Computing RUL values with a cap of 130 cycles
- Feature selection (24 features: 3 operational settings, 21 sensor measurements)
- Min-Max feature normalization
- Train-validation split (80%-20%) by engine ID

2.2 Support Vector Regression Implementation

Our SVR implementation utilized the RBF kernel with parameters:

- $C = 1.0$ (regularization parameter)
- $\varepsilon = 0.1$ (epsilon-tube width)

The RBF kernel is defined as:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{1}$$

2.3 Multi-Layer Perceptron Implementation

2.3.1 Initial Implementation

- Architecture: Input(24) $\rightarrow$ Hidden(64) $\rightarrow$ Output(1)
- Activation: ReLU
- Optimization: AdamW (lr=0.001, weight_decay=0.01)
- Loss function: MSE

2.3.2 Advanced Implementation

To address convergence issues, we implemented several improvements:

1. **Enhanced Architecture**:
   - Multiple hidden layers [256, 128, 64]
   - Leaky ReLU activation
   - Batch normalization
2. **Target Normalization**:
   - Standardization of target values
3. **Regularization Techniques**:
   - Dropout (rate=0.2)
   - Reduced weight decay (0.0005)
   - Gradient clipping (threshold=5.0)
4. **Optimization Enhancements**:
   - Learning rate scheduling (cosine annealing)
   - Initial learning rate: 0.01

3 Results and Analysis

3.1 Performance Comparison

Model	RMSE	Time (s)
SVR (RBF)	19.36	4.28
Initial MLP	98.29	3.38
Advanced MLP	17.82	18.57

Table 1: Performance comparison

The initial MLP implementation exhibited poor performance with minimal convergence, as evidenced by negligible changes in training loss:

```
Epoch 1/100, Train Loss: 9783.3101, Val Loss: 9660.3653
Epoch 100/100, Train Loss: 9783.2961, Val Loss: 9660.3572
```

In contrast, the advanced MLP demonstrated significant improvement, slightly outperforming SVR in prediction accuracy while requiring more training time.

3.2 Impact of Individual Improvements

Our ablation analysis revealed the relative importance of each improvement:

1. **Target Normalization**: Most critical improvement (57% reduction in RMSE)
2. **Architectural Changes**: Enhanced model capacity (25% further reduction)
3. **Batch Normalization**: Addressed internal covariate shift (23% further reduction)
4. **Learning Rate Scheduling**: Improved optimization dynamics (18% further reduction)
5. **Regularization Techniques**: Enhanced generalization (10% further reduction)

4 Discussion

4.1 SVR vs. MLP Comparison

While both models achieved comparable prediction accuracy, they exhibited different characteristics:

- **Implementation Complexity**: SVR required minimal tuning, while MLP demanded extensive optimization adjustments.
- **Computational Efficiency**: SVR demonstrated superior training efficiency (4.28s vs. 18.57s).