Performance of Hypothesis Classes on Time Series Data

Sasmit Datta Debangshu Chowdhary Netra Poonia

1 Problem Statement

This report aims to evaluate the predictive capabilities of three distinct models: a linear model, a decision tree, and a neural network, to determine which of them provides the most accurate and reliable forecasts for a given time series dataset. Each model comes with its own strengths and weaknesses concerning time series data - linear models are simple and interpretable but may fail to capture complex patterns; decision trees offer a more nuanced approach to non-linearity but may overfit; and neural networks offers a parametrized way to fit non-linear data but can suffer from high computational demands.

1.1 Formulation

For given time series data $\mathbf{x}_{T-k}, \mathbf{x}_{T-k+1}, \ldots, \mathbf{x}_T$ where $\mathbf{x}_t \in \mathbb{R}^n$ we fit a model

$$f(\mathbf{x}_{T-k}, \mathbf{x}_{T-k+1}, ..., \mathbf{x}_{T-1}) = \mathbf{x}_T^{(n)} \tag{1}$$

where $\mathbf{x}_t^{(i)} \in \mathbb{R}$ is the i-th value of the vector $\mathbf{x}_t$. In essence what we are trying to do is given the multivariate time-series data points of the k previous time steps, we try to predict the scaler value of the n-th index of data point of the next time-step.

2 Methodology

2.1 Dataset