THERMAL MODELING
FOR AVERAGE SIZE COMMERCIAL BUILDINGS

Aishwarya Mukunda

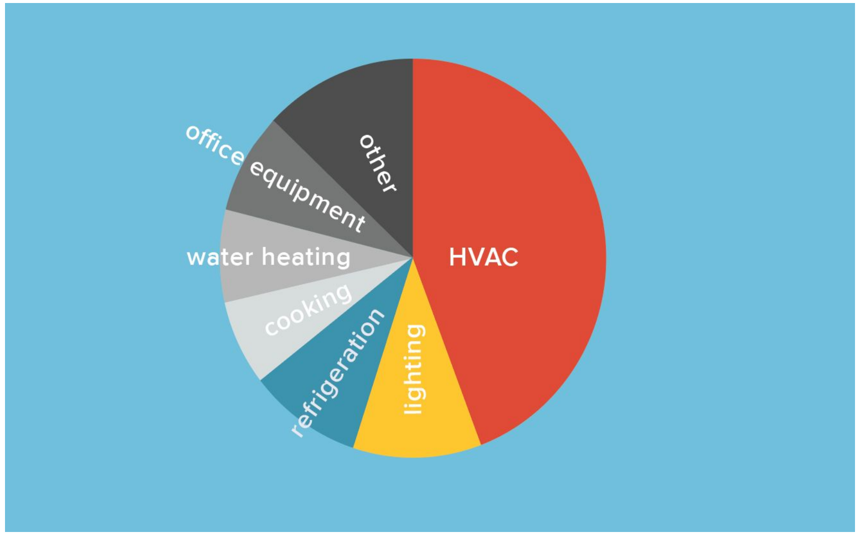
Advisor: Athanasios Aris Panagopoulos

OUTLINE

- Motivation
- Research Objective
- Related Work & Background Material
- Our Approach
- Implementation
- Results
- Conclusion
- Future Work

MOTIVATION

- In USA, commercial buildings consume 70% of US electricity.
- They spend lot of money and energy in Heating, Ventilating and Air Conditioning (HVAC) systems.
- Intelligent applications and systems can reduce cost.
- They need thermal models.
- Thermal model – predicts thermostat temperature based on several factors such as indoor temperature, outside temperature, occupancy and humidity.



RESEARCH OBJECTIVE

Current thermal models are application specific, simplistic and/or difficult to generalize.

The aim of this research project is to build a thermal model that meets the following criteria:

- Generality
- efficient
- handle with intermittent and missing data
- robust and consistent
- trained in real operation time
- requires minimum, and readily available knowledge about the building

RELATED WORK & BACKGROUND

THERMAL MODELING

- White box
  - Require extensive knowledge about the building structure.
- Black box
  - They do no generalize untrained regions.
  - No physical meaning.
- Gray box
  - Combines these two

GRAY BOX THERMAL MODELING

- For instance, simplified thermal equations can be defined as

$$T(t+1) = c * T(t)$$

  where c represents collectively the thermal dynamics

- Data
- Supervised learning
- Statistical techniques

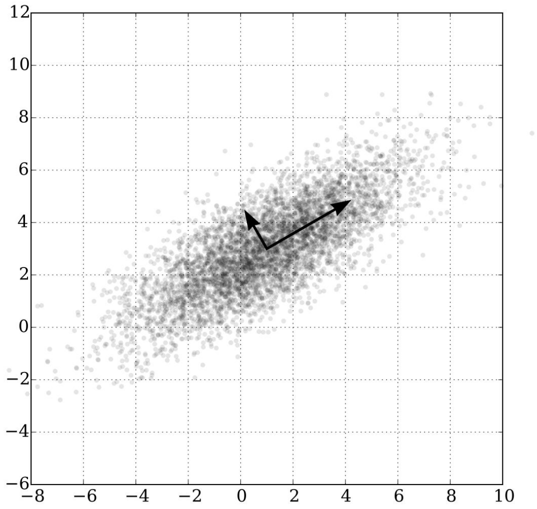
SUPERVISED LEARNING

- Data preprocessing
- Feature identification
- Training
- Evaluation

- Select an appropriate model

PRINCIPAL COMPONENT ANALYSIS

- Transforms high dimensional vector space into a low dimensional space.
- Improves performance without losing important information.
- It finds the direction of maximum variance (first principal component)
- It finds the direction of next maximum variance (second principal component) such that it is orthogonal to the first principal component.



PRINCIPAL COMPONENT ANALYSIS IN THERMAL MODELLING

- *"Thermal Models for Intelligent Heating of Buildings"*
- Thavlov, Anders and Henrik
- It was used to process the collected temperature data.
- Reduced the dimensionality from eight measurements to one parameter representing aggregated temperature.
- This proved better than averaging the temperature when plotted on a graph.
- Hence, we use the same method to reduce the dimensionality of our data.

LINEAR REGRESSION

Make predictions of continuous values.

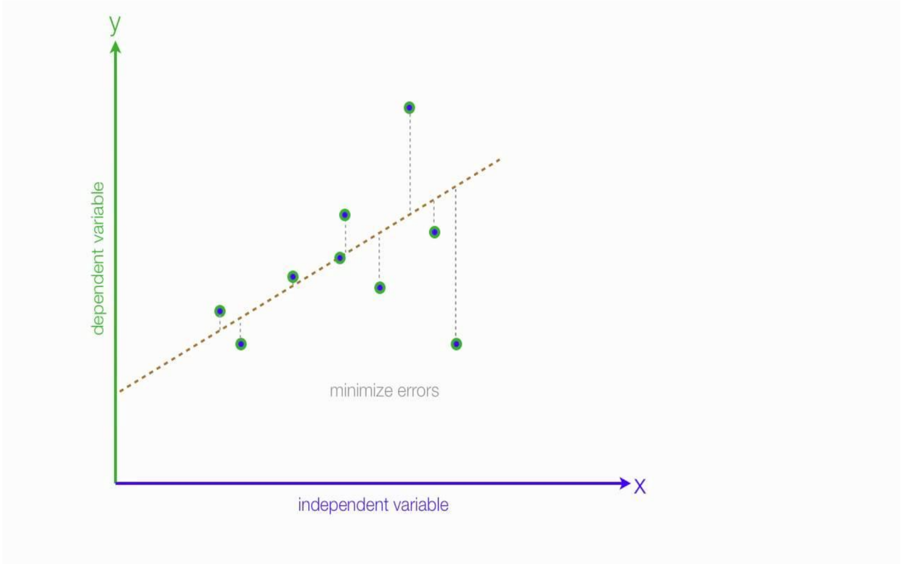
Linear relationship between the predicted variable and the variables that determines the predicted variable.

Regression line is defined as,

$$\hat{y} = w[0] \times x[0] + w[1] \times x[1] + \dots + w[n] \times x[n] + b$$

Cost function:

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{p} w_j \times x_{ij} \right)^2$$



LINEAR REGRESSION IN THERMAL MODELING

"Linear regression models for prediction of annual heating and cooling demand"

Navid Aghdaie, Georgios Kokogiannakis, Daniel daly, Timothy McCarthy

The total annual heating and cooling energy requirements was predicted using multiple linear regression model.

It is easier and more practical to implement linear model when compared to non-linear.

Many optimizations rely on linear programming

The results of this paper indicates the linear dependence between the predicted variable and its features.

LASSO REGRESSION

- The cost function of linear regression is further regularized by including magnitude of weights (L1 regularization).

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \boxed{\lambda \sum_{j=0}^{p} |w_j|}$$

- Regularization parameter penalizes the weights to get a simpler model.
- This can lead to zero weights which makes the features neglected.
- Hence it helps in both feature selection.

LASSO REGRESSION IN THERMAL MODELING

- “*Machine Learning-Based Temperature Prediction for Runtime Thermal Management Across System Components*”
- Zhang Kaicheng, Guliani Akhil
- This method was adapted because it performs :
  - Feature selection
  - Penalizes the feature weights, if the feature causes errors in training set.

N-ORDER LINEAR MODELS

- Time series autocorrelation

First Order: $$T(t+1) = c_1 * T(t)$$

Second Order: $$T(t+1) = c_1 * T(t) + c_2 * T(t-1)$$

Third Order: $$T(t+1) = c_1 * T(t) + c_2 * T(t-1) + c_3 * T(t-2)$$

.

.

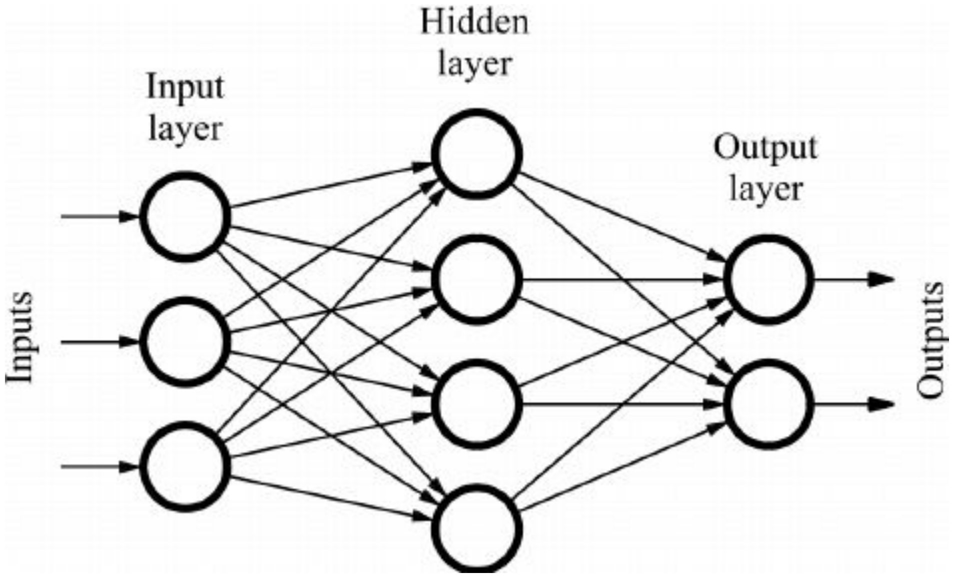
N Oder:

"*Modeling Multiple Time series with applications*"

- G.C Tiao $$T(t+1) = c_1 * T(t) + c_2 * T(t-1) + c_3 * T(t-2) + \ldots + c_n * T(t-n)$$
- The order of the model indicates the number of previous timesteps considered to predict the future time.

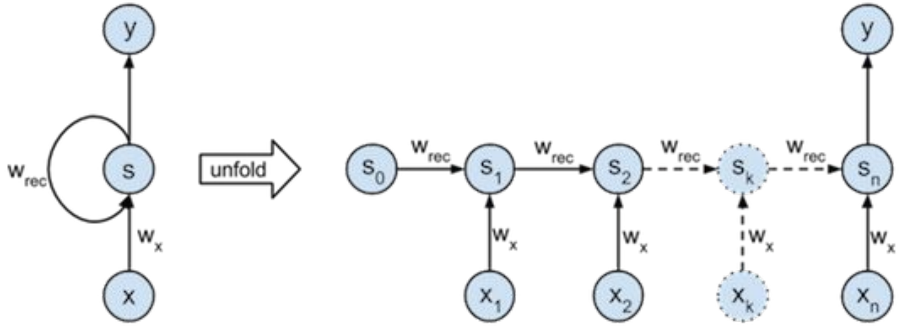
NEURAL NETWORKS

- Feed forward network
  - Information flows in one direction(input to output layer) only.
  - Hence less suitable for sequences or time series data



RECURRENT NEURAL NETWORKS

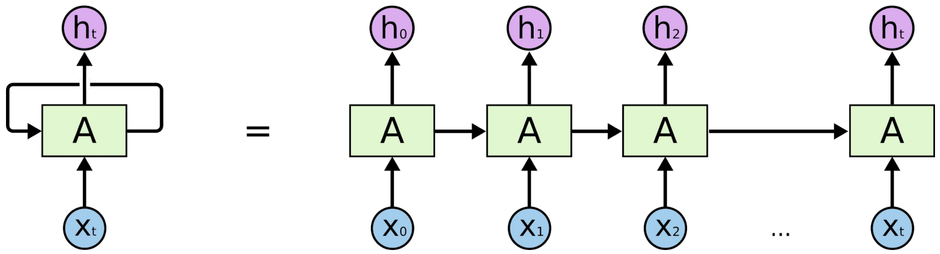
- Information flows in both directions.
- Suitable for time series data.
- Backpropagation through time
- Gradient(small number between 0 and 1) at a layer is product of gradients at prior layers.
- Hence in earlier layers, the weights are not changed significantly and learning does not happen. This is called vanishing gradient problem

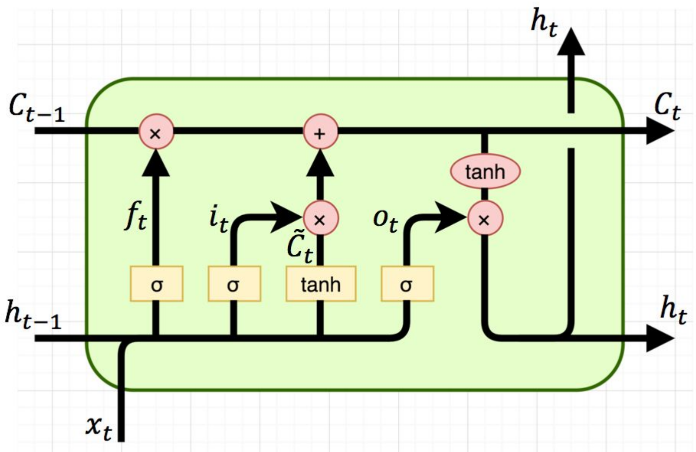


LONG SHORT TERM MEMORY(LSTM)

- Solves vanishing gradient problem
- Cell state: memory block that holds all relevant information
- 3 gates:
  - Forget: decides what information to remove from block
  - Input: decides values from input required to update memory
  - Output: decides the output based on input and memory

Since the first layer performs PCA, we do not apply PCA.





LSTM IN THERMAL MODELING

- *"Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm"*
- Salah Bouktif, Ali Fiaz, Ali Ouni and Mohamed Adel Serhani
- Predictions for Electric load is made for 2 days to mid weeks.
- They used LSTM, a well-suited model for forecasting time series data.
- It overcomes the problem of vanishing gradients in recurrent neural networks.
- It can pass past inputs into the network and also predict a sequenced output.

OUR APPROACH

OUR APPROACH

Methods Used:

- Principal Component Analysis
- Linear Regression
- Lasso Regression
- LSTM

Each method is applied to :

- a simple model (first order, containing features of current timestamp)
- a complex model (second order, containing features of current and previous timestamp)

PREPROCESSING

Convert time-series data to supervised learning problem.

Time(t)	Temperature(t)

Time(t)	Temperature(t)	Thermostat Temperature(t+delta)

One-Hot Encoding: Convert multi-category feature to binary features.

Time Index(t)	Action
t1	No action
t2	heating
t3	cooling

TimeIndex	Action_heating	Action_cooling
t1	0	0
t2	1	0
t3	0	1

PREPROCESSING

Quality control test: Unrealistic temperature data are removed by consulting the experts at the UC Berkeley.

Standardization:

It is a common technique to convert the dataset into common scale so that all features are comparable.

It is necessary for PCA.

Standard Scaler:

Convert the values of features such that the mean=0 and variance=1

Standard score, z = (x – mean) / standard_deviation

LINEAR REGRESSION

- Dependent variable: Temperature at next delta minutes
- Independent variables:
  - Indoor Temperature
  - Outdoor Temperature
  - Action (heating, cooling, no action)
  - Occupancy (binary)

LINEAR REGRESSION

- 1^{st} order:

$$T\_in(t+delta) = c_1 * t\_in(t) + c_2 * t\_out(t) + c_3 * occ(t) +$$
$$c_4 * action\_heating(t) + c_5 * action\_cooling(t) + intercept$$

- 2^{nd} order:

$$T\_in(t+delta) = c_1 * t\_in(t) + c_2 * t\_out(t) + c_3 * occ(t) + c_4 * action\_heating(t)$$
$$c_5 * action\_cooling(t-1) + c_{11} * t\_in(t-1) + c_{12} * t\_out(t-1) + c_{13} * occ(t-1) +$$
$$c_{14} * action\_heating(t-1) + c_{15} * action\_cooling(t-1) + intercept$$

LINEAR REGRESSION WITH PCA

PCA: N dimensions are reduced to k dimensions, while maintaining high variance.

f1	f2	…	fn

f1	…	fk

- 1st order:
- 2nd $T\_in(t + delta) = c_1 * x\_1(t) + c_2 * x\_2(t) + \ldots + \boxed{c_k * x\_k} + intercept$

$$T\_in(t + delta) = c_1 * x\_1(t) + c_2 * x\_2(t) + \ldots + c_k * x\_k + c_{21} * x\_1(t-1) + c_{22} * x\_2(t-1) + \ldots + c_{2k} * x\_k(t-1) + intercept$$

LSTM

- 3 layers:

Name of Layer	Number of Layers	Number of neurons
Input	1	Number of features
Output	1	Number of outputs
Hidden	1	Mean of number of neurons in input and output layer

After trying with different architectures we used the rule of thumb.

CASE STUDY

CASE STUDY

- Evaluate in 3 stores
- Perform separate training and testing
- Evaluate by calculating the Root mean squared error (rmse) and Mean squared error (mse)

DATASET

The data is collected from Corporation Yard at Berkeley. It comprises of the following stores:

- Electrical shop
- Radio shop
- FCA Main Department

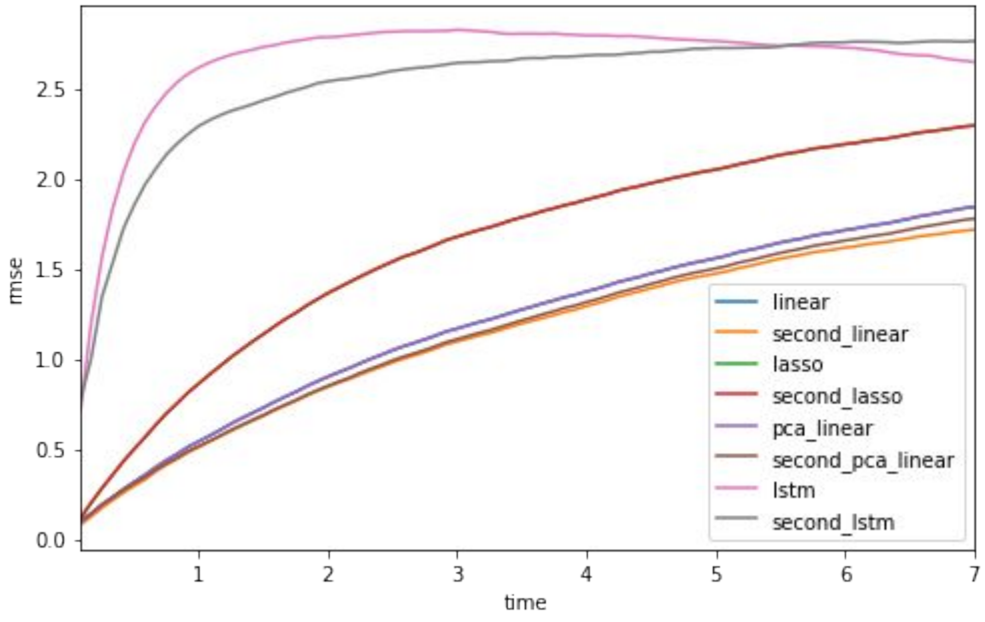
City of Berkeley – Corporation Yard

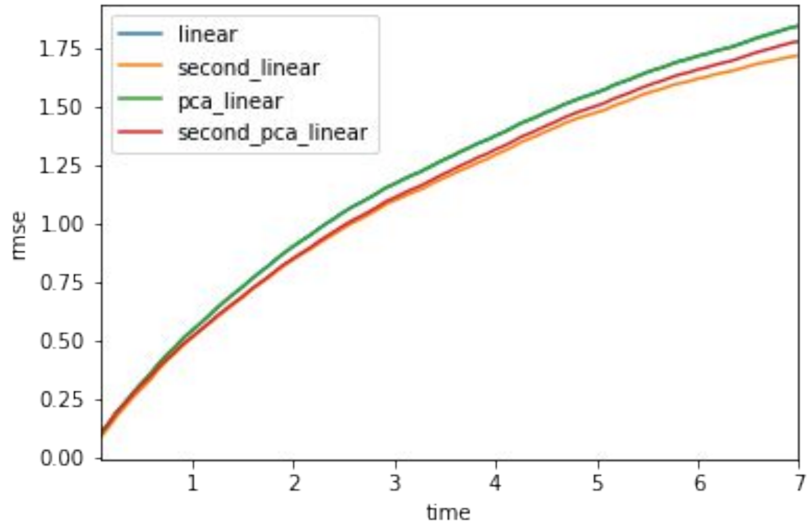
The dataset contains the following features:

- Historic and Current Indoor Temperature
- Historic and current Outdoor Temperature data, aggregated from multiple nearby weather stations.
- Electric consumption
- Historic Heating and Cooling Actions
- Occupancy Schedule

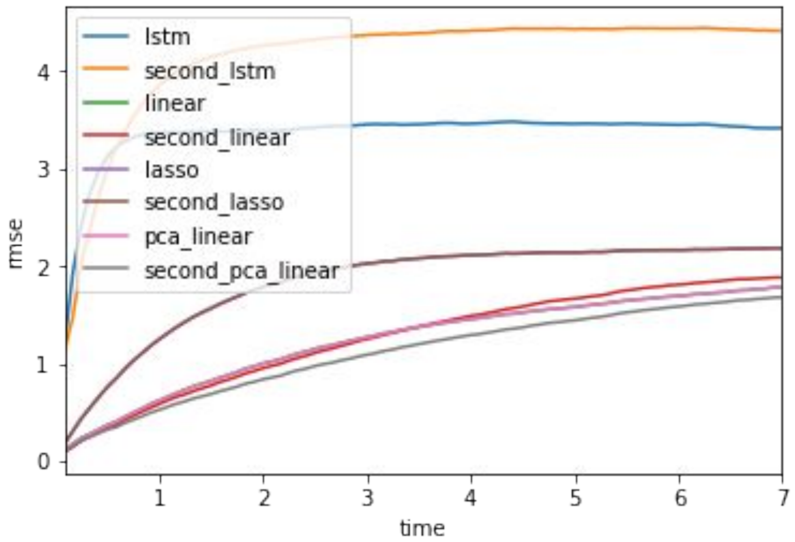
THE RESULTS

Comparing models with respect to forecast prediction.

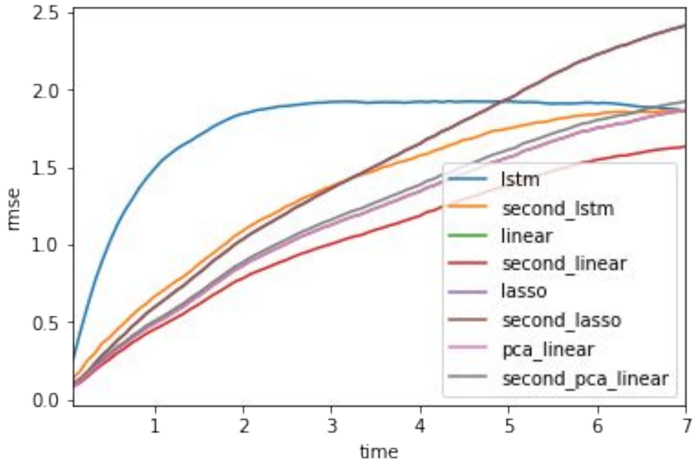




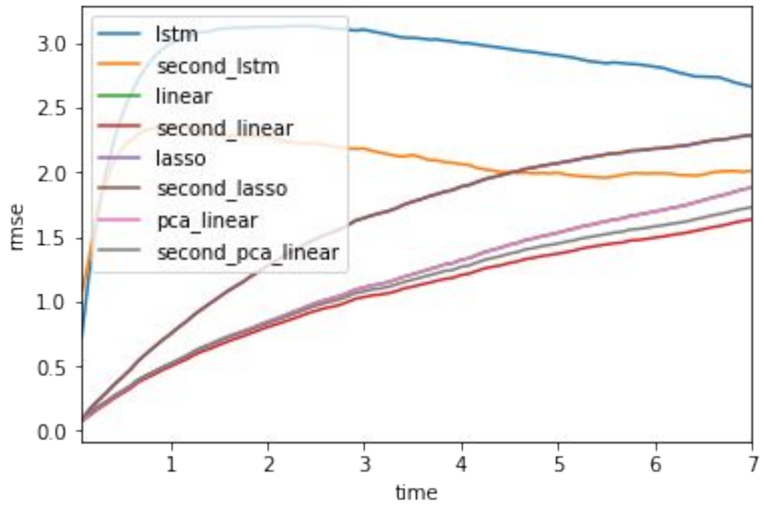
Electrical Shop



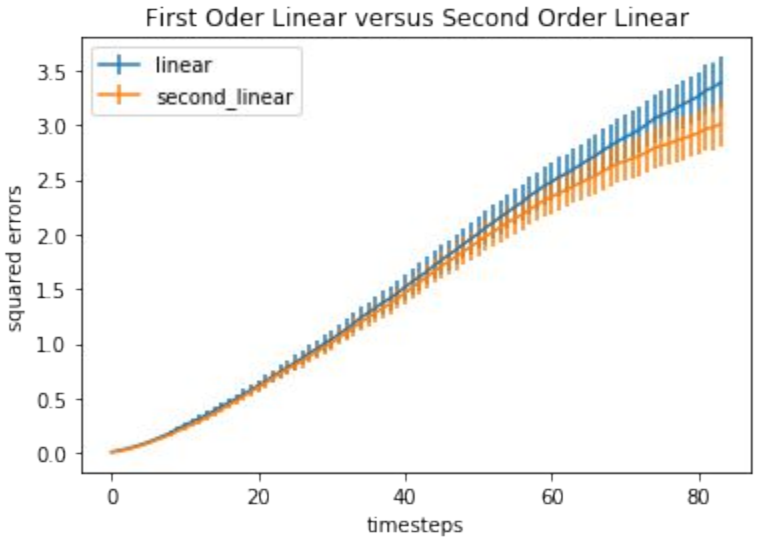
Radio Shop

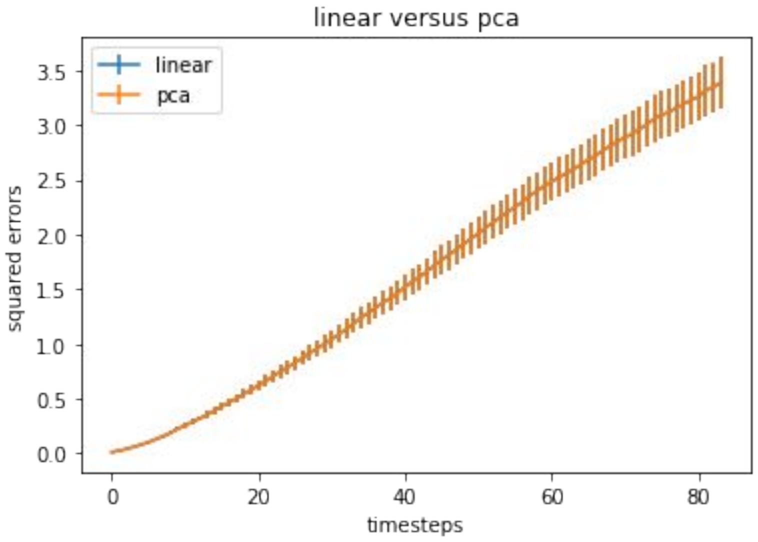


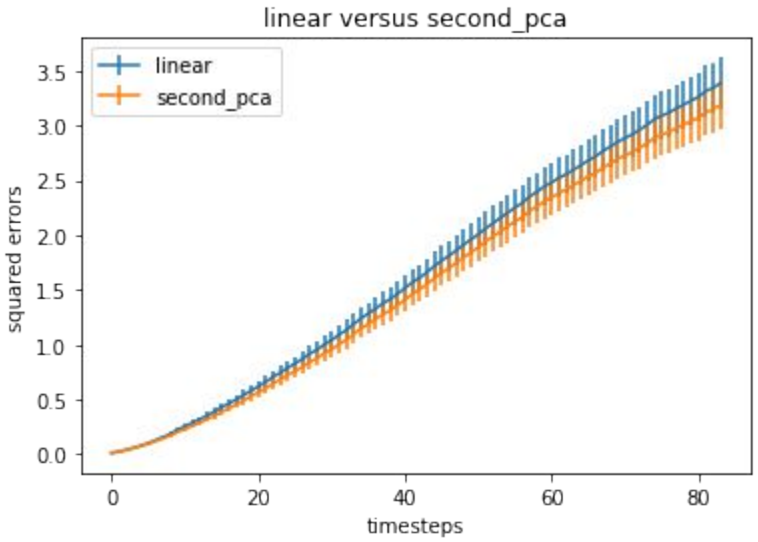
FAC Main Dept

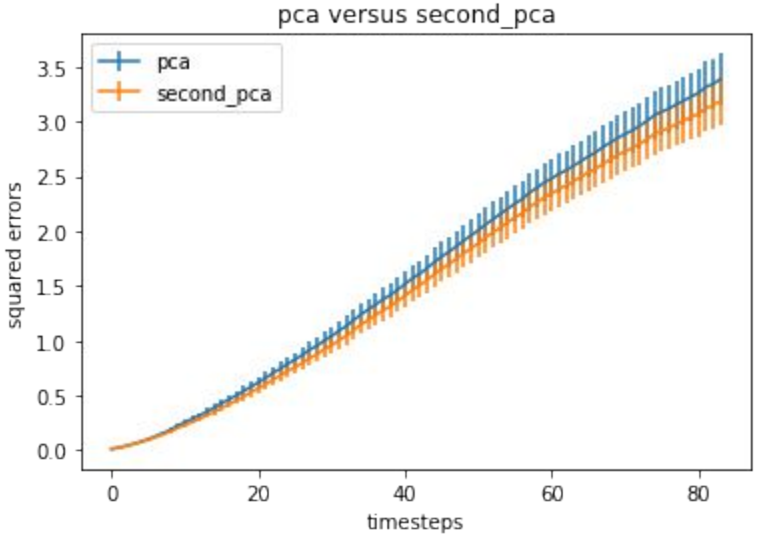


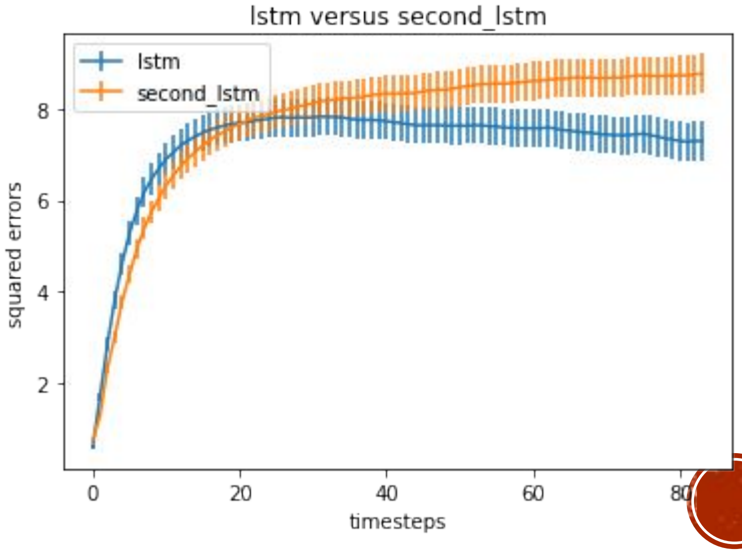
GENERAL MODEL











T-TEST

T-Test is used to determine if there is a significant difference in the mean between two populations or the difference observed is by a random chance.

Pairwise t-test: Apply two models to same test data and compare their error.

Null Hypotheses: The mean of the difference of the means is zero.

Alternate Hypotheses: There is a difference in mean between the two given models.

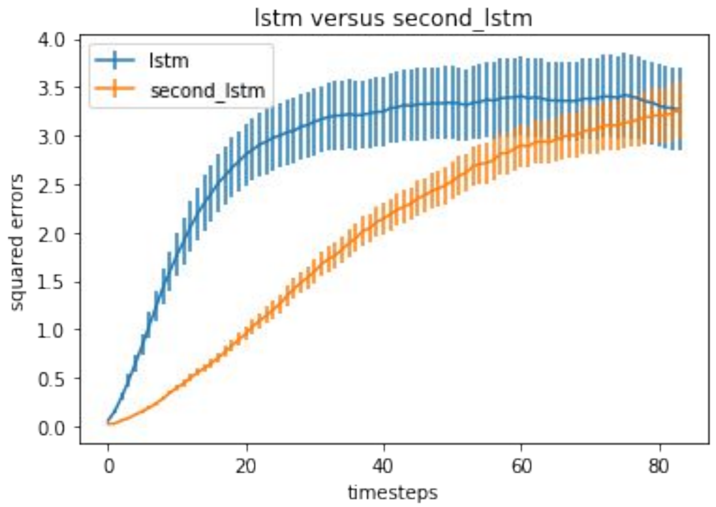
If p-value is smaller than 0.05, then reject the null hypotheses.

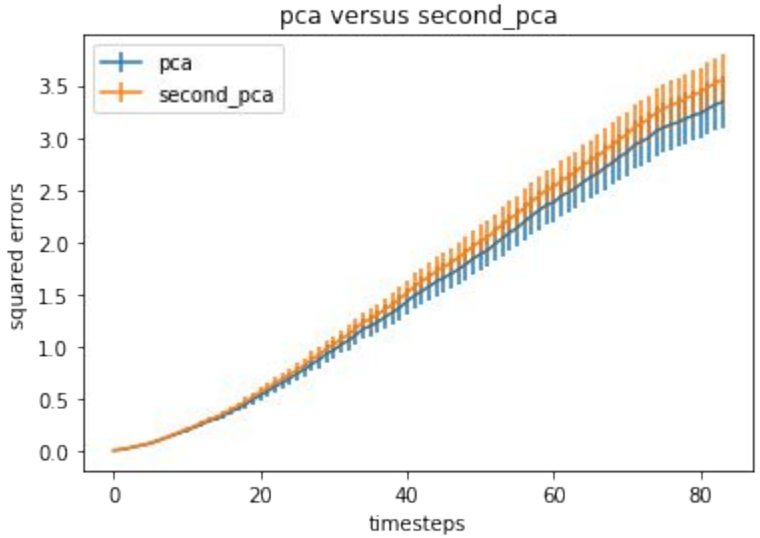
there exists a statistical significance in difference between the two models

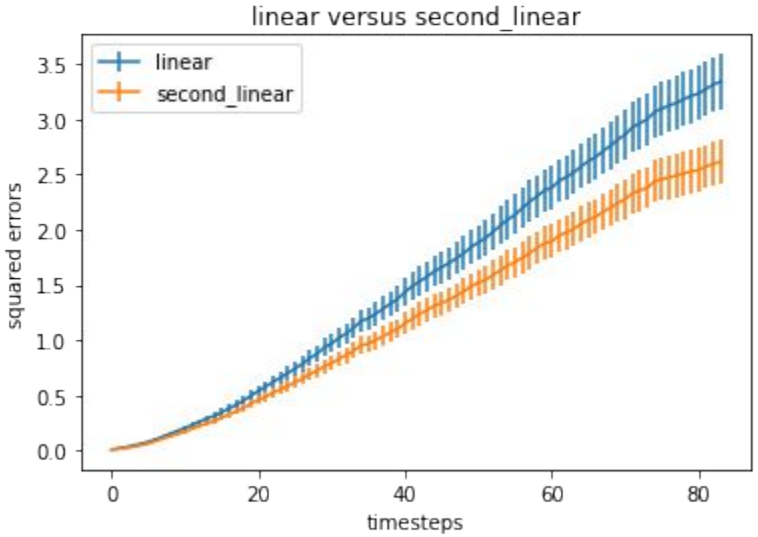
T- Test Results

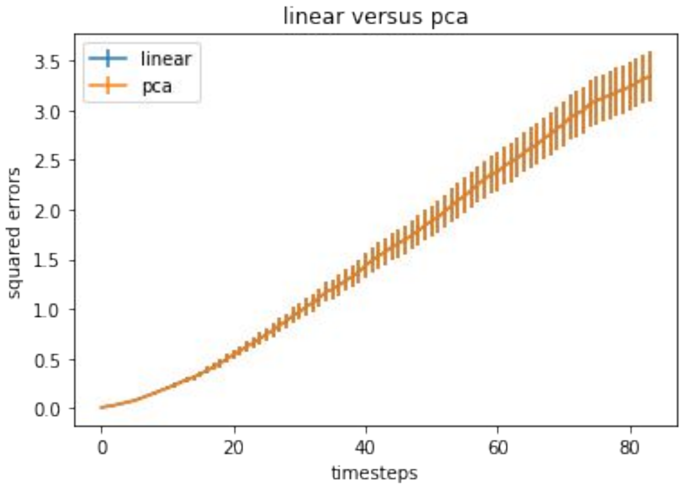
Significant	Insignificant	Both
Linear and LSTM (both first and second order combinations)	First Order Linear and First Order Linear with PCA	First and Second Order Linear
First Order Linear with PCA and Second Order Linear with PCA		Second Order Linear and PCA
		First Order LSTM and Second Order LSTM

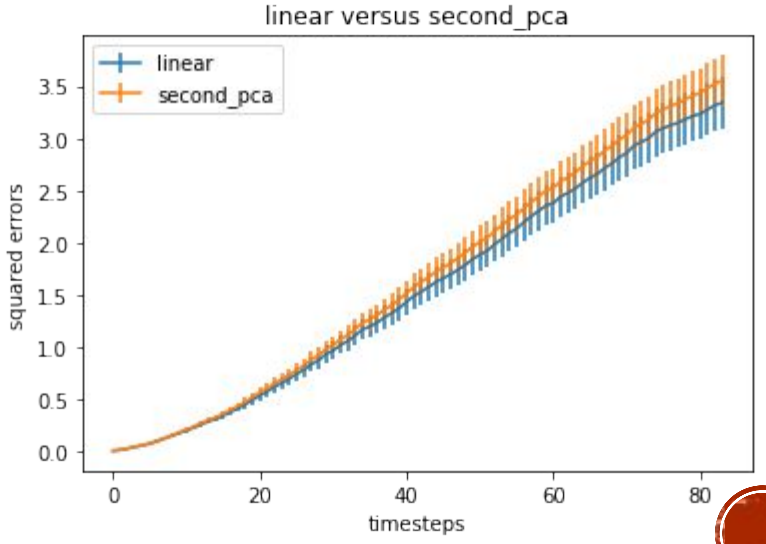
Radio Shop


lstm versus second_lstm
lstm
second_lstm
squared errors
timesteps

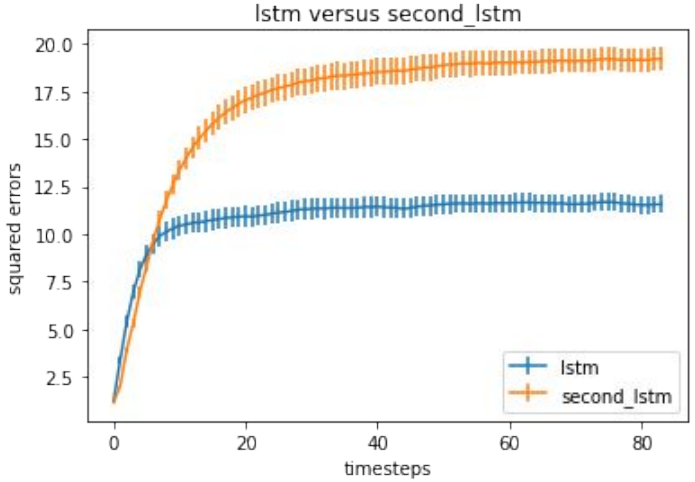

pca versus second_pca
pca
second_pca
squared errors
timesteps

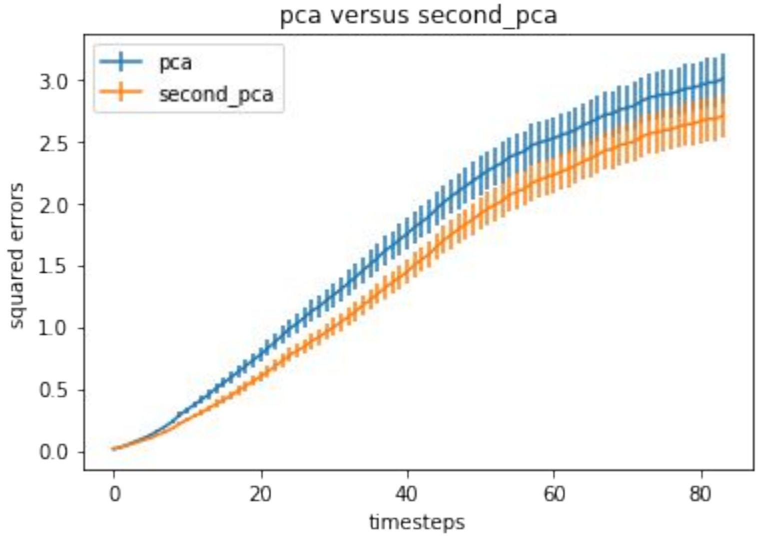

linear versus second_linear
linear
second_linear
squared errors
timesteps

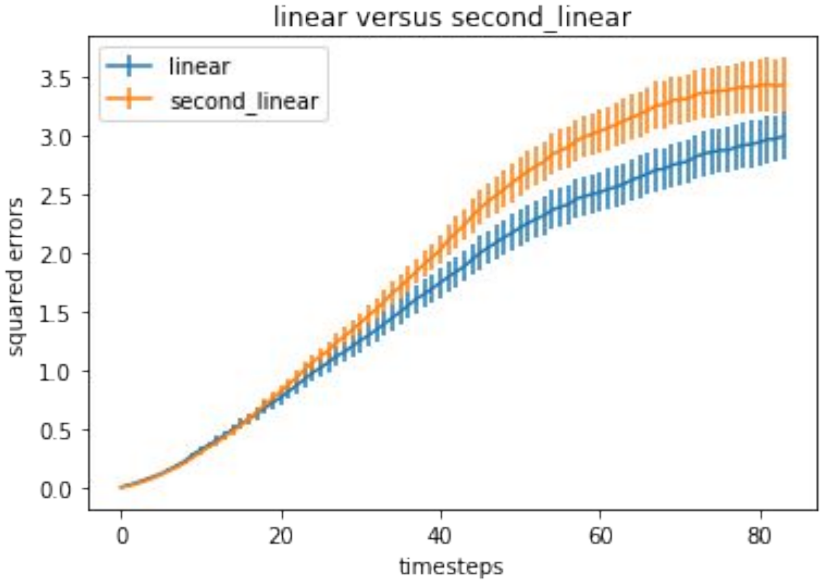

linear versus pca
linear
pca
squared errors
timesteps

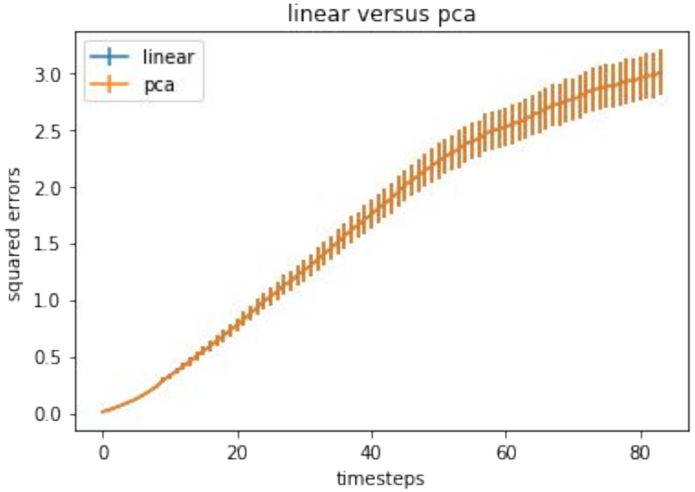

linear versus second_pca
linear
second_pca
squared errors
timesteps

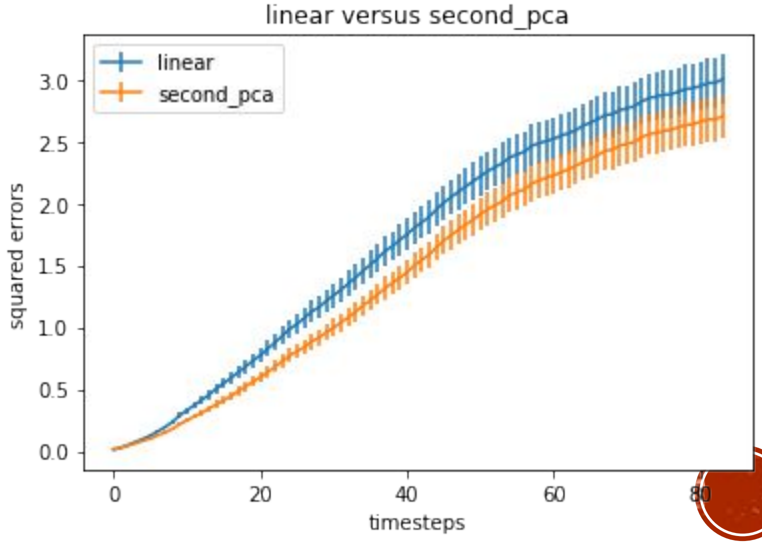
Electrical Shop



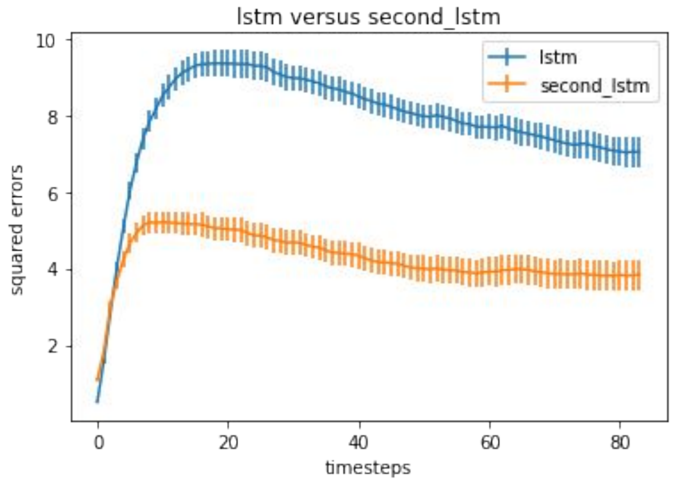




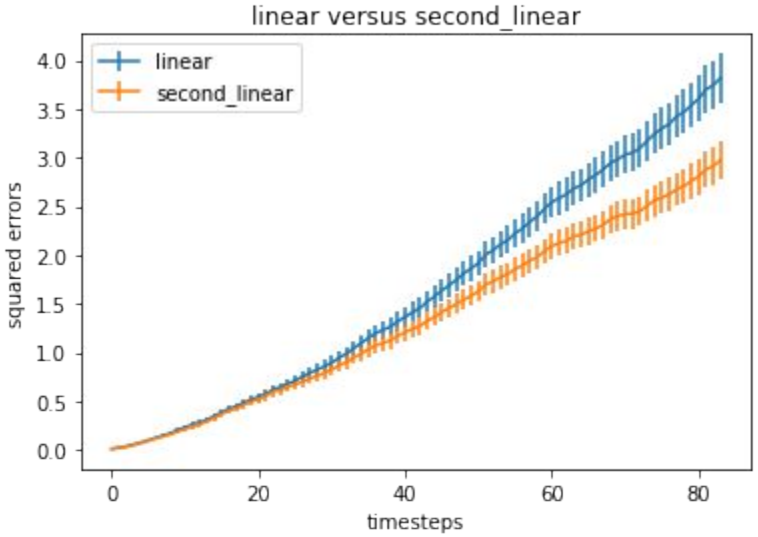




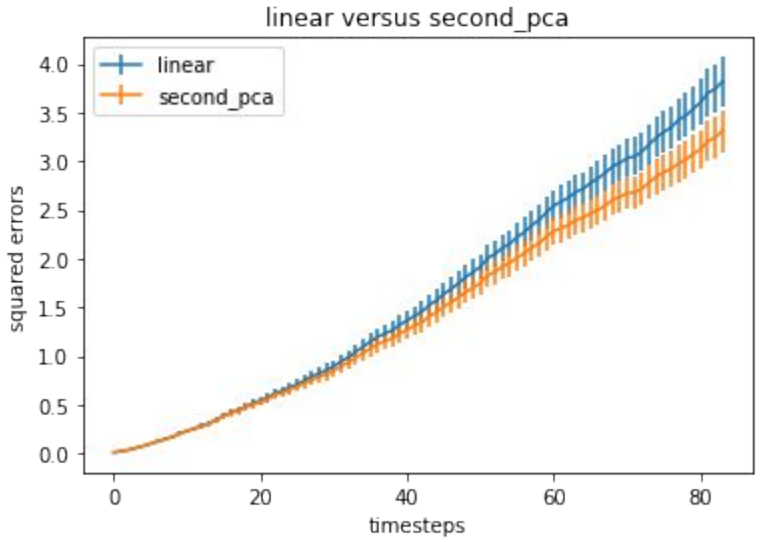
FAC



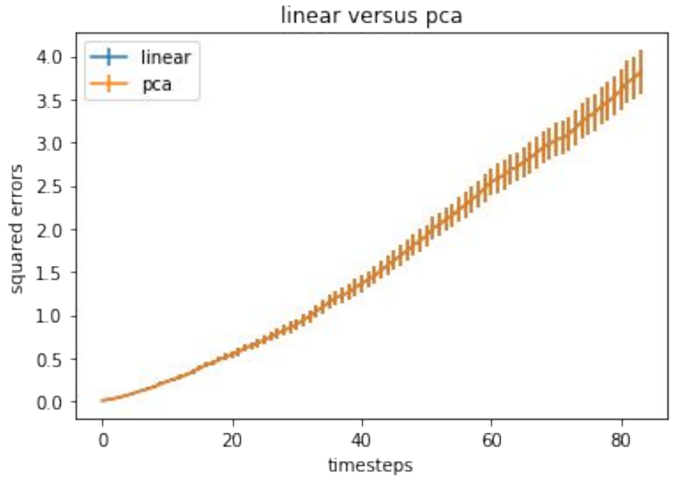
lstm versus second_lstm
lstm
second_lstm
squared errors
timesteps



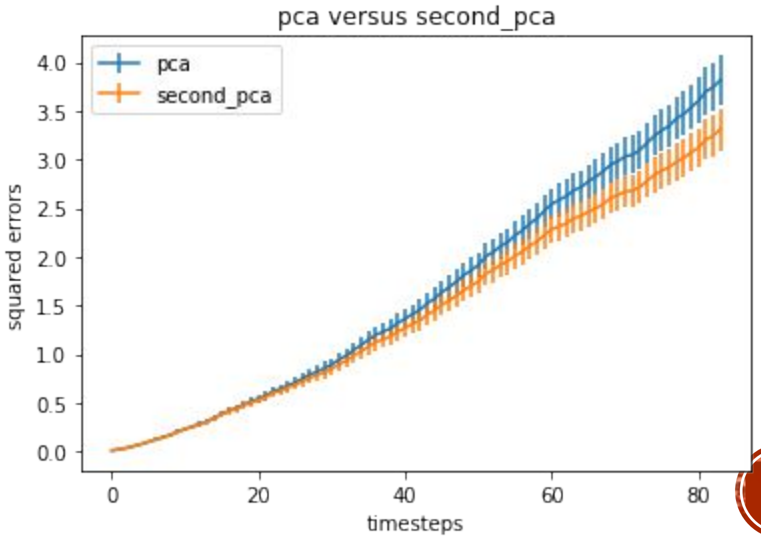
linear versus second_linear
linear
second_linear
squared errors
timesteps



linear versus second_pca
linear
second_pca
squared errors
timesteps



linear versus pca
linear
pca
squared errors
timesteps



pca versus second_pca
pca
second_pca
squared errors
timesteps

CONCLUSION

- Dataset was collected from the existing information available and did not require extra instrumentation.
- Dataset is preprocessed to take care of gaps in time intervals and missing data.
- PCA and Lasso did not help much as we used only the important features.
- Surprisingly, LSTMs did not give better results. As by the rule "Keep it simple", linear regression works best.
- First order and second order linear model showed almost similar results, hence on the grounds of simplicity we choose first order linear model.
- Linear model fits well for the above buildings. Hence, accounting for generality.
- Linear model gives less error rate, hence proving efficiency.
- Predictions were made in real time and hence accountable for real-time operation.

FUTURE WORK

- Investigate further with other Machine Learning approaches
- Incorporate the thermal model in intelligent applications.
- Adaptive thermal modeling

THANK YOU