Application Development Laboratory (CS 33002)

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

School of Computer Engineering

2 Credit

Analytics Application Development using R

Lab Contents

Sr #	Major and Detailed Coverage Area	Lab#
1	MSE, RMSE and MAE	10
2	Time Series Forecasting	

Performance Analysis

- What makes a good prediction? Of course, a good prediction is an accurate prediction.
- A prediction "error" is the difference between an observed value and its prediction. The "error" does not mean a mistake, it means the unpredictable part of an observation.
- Error measure plays an important role in calibrating and refining prediction model/method and helps the analyst to improve prediction method.
- The choice of an error measure may vary according to the situation.
- The popular and highly recommended error measures are
  - Mean Square Error (MSE)
  - Root Mean Square Error (RMSE)
  - Mean Absolute Percentage Error (MAPE)

Mean Square Error (MSE)

MSE is defined as mean or average of the square of the difference between actual and estimated values. Mathematically it is represented as:

$$\text{MSE} = \frac{\sum_{j=1}^{N} (\text{observation } (j) - \text{prediction } (j))^2}{N}$$

Month	1	2	3	4	5	6	7	8	9	10	11	12
Actual Demand	42	45	49	55	57	60	62	58	54	50	44	40
Forecasted Demand	44	46	48	50	55	60	64	60	53	48	42	38
Error	-2	-1	1	5	2	0	-2	-2	1	2	2	2
Squared Error	4	1	1	25	4	0	4	4	1	4	4	4

Sum of Square Error = 56 and MSE = 56 / 12 = 4.6667

MSE Calculation Example


```
data <- data.frame(actual=c(35, 36, 43, 47, 48, 49, 46, 43, 42, 37, 36, 40),
predicted=c(37, 37, 43, 46, 46, 50, 45, 44, 43, 41, 32, 42))
print(data)
```

Computing MSE:

The below code calculates the MSE between the actual and predicted values by using the MSE function:

```
print(mse(data$actual, data$predicted))
```

Root Mean Square Error (RMSE)

It is just the square root of the mean square error. Mathematically it is represented as:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N} (\text{observation } (j) - \text{prediction } (j))^2}{N}}$$

Month	1	2	3	4	5	6	7	8	9	10	11	12
Actual Demand	42	45	49	55	57	60	62	58	54	50	44	40
Forecasted Demand	44	46	48	50	55	60	64	60	53	48	42	38
Error	-2	-1	1	5	2	0	-2	-2	1	2	2	2
Squared Error	4	1	1	25	4	0	4	4	1	4	4	4

Sum of Square Error = 56, MSE = 56 / 12 = 4.6667, RMSE = SQRT(4.667) = 2.2

RMSE Calculation Example


```
data <- data.frame(actual=c(35, 36, 43, 47, 48, 49, 46, 43, 42, 37, 36, 40),
predicted=c(37, 37, 43, 46, 46, 50, 45, 44, 43, 41, 32, 42))
print(data)
```

Computing RMSE:

Method 1: The below code calculates the RMSE between the actual and predicted values manually by following the RMSE formula:

```
print(sqrt(mean((data$actual - data$predicted)^2)))
```

Method 2: The Metrics package offers a convenient rmse() function. First, install and load the package:

```
install.packages("Metrics")
library(Metrics)

# Calculate RMSE using the rmse() function
rmse_value <- rmse(data$actual, data$predicted)
print(rmse_value)
```

Mean Absolute Percentage Error (MAPE)

The formula to calculate MAPE is as follows:

$$\text{MAPE} = (100/n) \times \sum_{i=1}^{n} \frac{(|X'(t) - X(t)|}{X(t)}$$

Here, X'(t) represents the prediction data value of point t and X(t) represents the actual data value of point t. Calculate MAPE for the below dataset.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Actual Demand	42	45	49	55	57	60	62	58	54	50	44	40
Forecasted Demand	44	46	48	50	55	60	64	60	53	48	42	38

- MAPE is commonly used because it's easy to interpret and easy to explain. For example, a MAPE value of 11.5% means that the average difference between the forecasted value and the actual value is 11.5%.
- The lower the value for MAPE, the better a model is able to forecast values e.g. a model with a MAPE of 2% is more accurate than a model with a MAPE of 10%.

MAPE Calculation Example


```
data <- data.frame(actual=c(35, 36, 43, 47, 48, 49, 46, 43, 42, 37, 36, 40),
predicted=c(37, 37, 43, 46, 46, 50, 45, 44, 43, 41, 32, 42))
print(data)
```

Computing MAPE:

Method 1: The below code calculates the MAPE between the actual and predicted values manually by following the MAPE formula:

```
print(mean(abs((data$actual-data$forecast)/data$actual)) * 100)
```

Method 2: The MLmetrics package offers a convenient MAPE() function. First, install and load the package:

```
install.packages(" MLmetrics ")
library(MLmetrics)

# Calculate RMSE using the rmse() function
mape_value <- MAPE(data$actual, data$predicted)
print(mape_value)
```

Class Exercise

Calculate Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Student #	Number of hrs	Observed Score	Predicted Score
1	1	68	71
2	1	76	75
3	2	74	72
4	3	80	83
5	4	76	79
6	5	78	77
7	6	81	83
8	4	84	81
9	5	86	83
10	6	83	85

Time Series Forecasting

https://www.pluralsight.com/guides/time-series-forecasting-using-r

Please refer to above site. It explains Time Series Forecasting using R. Also there is a problem statement on unemployment. You can download the data and follow the step-by-step process. Subsequently, implement:

- Simple Exponential Smoothing
- Holt's Trend Method
- ARIMA
- TBATS

Thank You
End of Lab 7

Lab Experiments

1. Search and download at least 2 dataset related to linear regression. WAP to demonstrate RMSE, MSE and MAE.
2. Search and download at least 2 dataset related to multiple regression. WAP to demonstrate RMSE, MSE and MAE.
3. Search and download a dataset related to time series forecasting. WAP to demonstrate
   - Simple Exponential Smoothing
   - Holt's Trend Method
   - ARIMA
   - TBATS