Model Evaluation

INSAID

Course Overview

You are here...

Term	CDF	GCD	GCDAI	PGPDSAI
Term 1	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python
Term 2	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques
Term 3	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling
		Minor Project	Minor Project	Minor Project
Term 4		Machine Learning Foundation	Machine Learning Foundation	Machine Learning Foundation
Term 5		Machine Learning Intermediate	Machine Learning Intermediate	Machine Learning Intermediate
Term 6		Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)
		Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)
		Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)
		Capstone Project	Capstone Project	Capstone Project
Term 7		Bonus: Industrial ML (ML – 4 & 5)	Basics of AI, TensorFlow, and Keras	Basics of AI, TensorFlow, and Keras
Term 8			Deep Learning Foundation	Deep Learning Foundation
Term 9			NPL – I/CV – I	CV – I
Term 10			NLP – II/CV – II	NLP – I
			Capstone Project	Capstone Project
Term 11				CV – II
Term 12				NLP – II
				NLP – III + CV – III
				AutoVision & AutoNLP
				Building AI product

Term Context

- Introduction to Machine Learning
- Linear Regression
- Logistic Regression
- **Model Evaluation Techniques** You are here…

INSAID

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- Adjusted R – Squared

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

Need of Model Evaluation

- **To estimate the performance of the model** for future instances.
- Whether the model will estimate good results or not.

Agenda

Need of Model Evaluation

Regression Measures

- **Mean Absolute Error (MAE)**
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- Adjusted R – Squared

Classification Measures

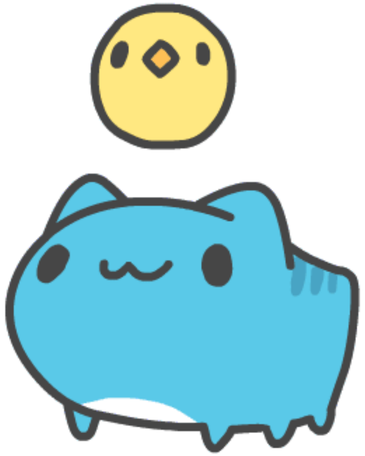
- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

1. Mean Absolute Error (MAE)

- It gives the **average difference** of actual and predicted values.
- They **don't give direction** of error i.e. **Under-Prediction** or **Over-Prediction**.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- y_i : Actual value
- $\hat{y}_i$: Predicted value
- n : No. of points



INSAID

Mean Absolute Error (MAE) Example

$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + c$

$y = 0.5 x_1 + 0.9 x_2 + 2.3 x_3 + 7$ (assume)

$$\frac{\sum_{i=1}^{n} |y_A - y_p|}{n}$$

Here, $n = 5$

	TV (X_1)	Radio (X_2)	Newspaper (X_3)	Sales(actual) (y)	Sales(predicted) (y_pred)
1	230 x_1	40 x_2	x_3 70	A1 22	P1 23.5
2	45	42	45	A2 10	P2 12
3	20	45	60	A3 15	P3 14.2
4	150	41	55	A4 20	P4 17.3
5	180	12	30	A5 25	P5 24.9

$= (|A_1 - P_1| + |A_2 - P_2| + |A_3 - P_3| - |A_4 - P_4| + |A_5 - P_5|) \div 5$

MAE = (|22 – 23.5| + |10 - 12| + |15 – 14.2| + |20 – 17.3| + |25 – 24.9|) ÷ 5

MAE = (1.5 + 2 + 0.8 + 2.7 + 0.1) ÷ 5

MAE = 7.1 ÷ 5

$\boxed{\text{MAE} = 1.42}$

INSAID

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- **Mean Squared Error (MSE)**
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- Adjusted R – Squared

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

2. Mean Squared Error (MSE)

- It gives the **average squared difference** of actual and predicted values.
- Also known as Mean Squared Deviation **(MSD).**
- It **incorporates** both **variance** (spread) and **bias** (average value from actual).

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (yi - \hat{y}_i)^2$$

- y_i : Actual value
- $\hat{y}_i$: Predicted value
- n : No. of points

Mean Squared Error (MSE) Example

			y	y-pred
TV	Radio	Newspaper	Sales(actual)	Sales(predicted)
230	40	70	22	23.5
45	42	45	10	12
20	45	60	15	14.2
150	41	55	20	17.3
180	12	30	25	24.9

$$= \frac{\sum_{i=1}^{n} (y_A - y_P)^2}{n}$$

MSE = ($(22 – 23.5)^2$ + $(10 – 12)^2$ + $(15 – 14.2)^2$ + $(20 – 17.3)^2$ + $(25 – 24.9)^2$) ÷ 5

MSE = (2.25 + 4 + 0.64 + 7.29 + 0.01) ÷ 5

MSE = 14.19 ÷ 5

MSE = 2.83

$RMSE = \sqrt{2.83}$

INSAID

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- **Root Mean Squared Error (RMSE)**
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- Adjusted R – Squared

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

3. Root Mean Squared Error (RMSE)

- It is the standard deviation of residuals (errors) or **spread** of **residuals**.
- It explains **concentration** of the data **around** the **line of best fit.**
- As compared to mean absolute error, **RMSE punishes large errors**.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

or $\sqrt{MSE}$

- y_i : Actual value
- $\hat{y}_i$: Predicted value
- n : No. of points

Root Mean Squared Error (RMSE) Example

TV	Radio	Newspaper	Sales(actual)	Sales(predicted)
230	40	70	22	23.5
45	42	45	10	12
20	45	60	15	14.2
150	41	55	20	17.3
180	12	30	25	24.9

$\text{RMSE} = (((22 - 23.5)^2 + (10 - 12)^2 + (15 - 14.2)^2 + (20 - 17.3)^2 + (25 - 24.9)^2) \div 5)^{1/2}$

$\text{RMSE} = ((2.25 + 4 + 0.64 + 7.29 + 0.01) \div 5)^{1/2}$

$\text{RMSE} = (14.19 \div 5)^{1/2}$

$\text{RMSE} = 1.68$

INSAID

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- **Mean Absolute Percentage Error(MAPE)**
- R – Squared
- Adjusted R – Squared

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

4. Mean Absolute Percentage Error(MAPE)

- Also known as **Mean Absolute Percentage Deviation (MAPD).**
- It usually **expresses accuracy** as a **percentage**.

$$= \frac{\sum_{i=1}^{n} \left| \frac{y_A - y_P}{y_A} \right|}{n} * 100$$

$$\text{MAPE} = \frac{1}{n} \Sigma \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100$$

- y_i : Actual value
- $\hat{y}_i$: Predicted value
- n : No. of points

INSAID

Mean Absolute Percentage Error(MAPE) Example

$$\frac{\Sigma \left| \frac{y_A - y_P}{y_A} \right| \times 100}{n}$$

TV	Radio	Newspaper	Sales(actual) y_A	Sales(predicted) y_P
230	40	70	22	23.5
45	42	45	10	12
20	45	60	15	14.2
150	41	55	20	17.3
180	12	30	25	24.9

$|y_A - y_P| \div |y_A|$

A = |22 – 23.5| ÷ |22| = 0.06

B = |10 – 12| ÷ |10| = 0.2

C = |15 – 14.2| ÷ |15| = 0.053

D = |20 – 17.3| ÷ |20| = 0.135

E = |25 – 24.9| ÷ |25| = 0.004

MAPE = [(A + B + C + D + E) ÷ 5] x 100

MAPE = [(0.06 + 0.2 + 0.053 + 0.135 + 0.004) ÷ 5] x 100

MAPE = [0.452 ÷ 5] x 100

MAPE = [0.0904] x 100

MAPE = 9.04 %

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- **R – Squared**
- Adjusted R – Squared

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

INSAID

5. R – Squared [0 - 1]

0.8 ↑ (Industry acceptable model)

- It explains **proportion** of the **variance** in the **dependent variable** that is predictable from the independent variable(s).
- Also known as **Coefficient of Determination** and value lies between **0 & 1** (higher, the better).
- **Shortcoming:** It **always increases with** addition of **more features**. i.e. IVs

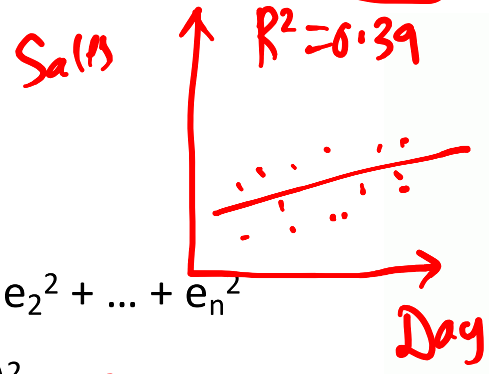
Dr. Darshan Ingle → Ph.D

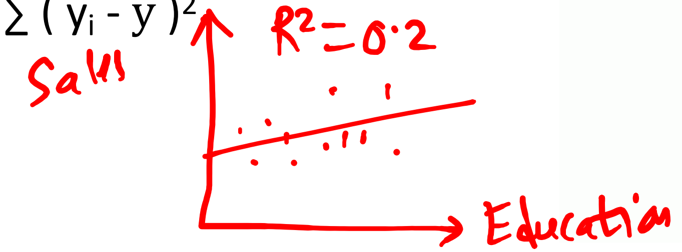
$$R^2 = 1 - \frac{RSS}{TSS}$$

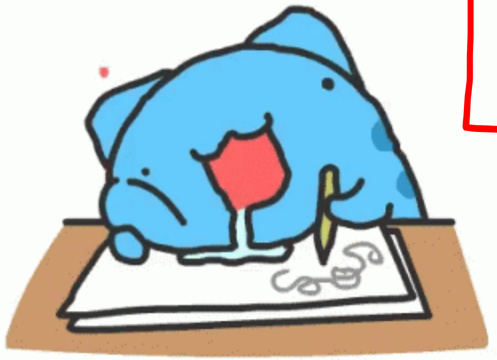
Residual Sum of Squares(RSS) = $e_1^2 + e_2^2 + \ldots + e_n^2$

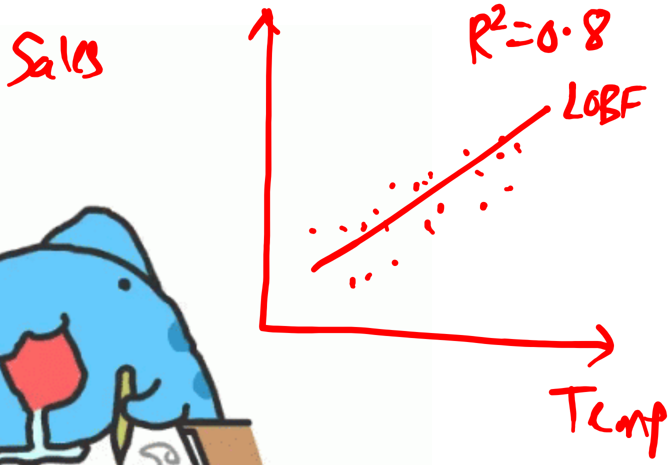
Total Sum of Squares(TSS) = $\sum (y_i - \bar{y})^2$

- y_i = Actual value
- $\bar{y}$ = Mean value
- e = Error









Temp is 80% responsible for predicting Sales

R – Squared Example

$R^2 = 1 - \left(\frac{RSS}{TSS}\right) = 1 - \left(\frac{e_1^2 + e_2^2 + \dots + e_n^2}{\Sigma(y_i - \bar{y})^2}\right)$

I.V. / Features / Predictors → X

D.V / Target variable → y

TV X_1	Radio X_2	Newspaper X_3	Sales(actual) y	Sales(predicted) y_pred
230	40	70	22	23.5
45	42	45	10	12
20	45	60	15	14.2
150	41	55	20	17.3
180	12	30	25	24.9

mean = 18.4 i.e, $\bar{y}$

R2

$X_1 \sim y \to 0.85$

$X_1 + X_2 \sim y \to 0.87$

$X_1 + X_2 + X_3 \sim y \to 0.91$

keeps growing

	R2
Temp ~ Sales	0.9
Temp + Education ~ Sales	0.94

* RSS = $(22 - 23.5)^2 + (10 - 12)^2 + (15 - 14.2)^2 + (20 - 17.3)^2 + (25 - 24.9)^2 = 14.19$
* TSS = $(22 - 18.4)^2 + (10 - 18.4)^2 + (15 - 18.4)^2 + (20 - 18.4)^2 + (25 - 18.4)^2 = 141.2$

$$R^2 = 1 - \frac{14.19}{141.2}$$

$R^2 = 1 - 0.100$ $\boxed{R^2 = 0.9}$

Agenda

Need of Model Evaluation

Regression Measures

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error(MAPE)
- R – Squared
- **Adjusted R – Squared**

Classification Measures

- Confusion Matrix
- Accuracy
- Classification Error
- Sensitivity / Specificity
- Precision / Recall
- F – Measure / Fβ – Measure
- ROC/AUC
- Classification Report

6. Adjusted R – Squared

Temp ~ Sales → A.R² 0.9

Temp+Educ · ~ Sales → 0.87

- It measures the **proportion** of **variance** explained by only those **predictors** that **really help** in **explaining** the **target** variable.
- It should be used to **compare models** of **different predictors** sizes or while performing **feature selection.**

$$\boxed{\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1}} =$$

$$1 - \left[\frac{(1-0.9)\,(5-1)}{5-3-1}\right] = 0.6$$

- n = No. of data points = 5
- p = No. of predictors i.e. no. of I.Vs = 3

INSAID

Adjusted R – Squared Example

TV	Radio	Newspaper	Sales(actual)	Sales(predicted)
230	40	70	22	23.5
45	42	45	10	12
20	45	60	15	14.2
150	41	55	20	17.3
180	12	30	25	24.9

$$\text{Adjusted } R^2 = 1 - \frac{(1-0.9)(5-1)}{5-3-1}$$

$$\text{Adjusted } R^2 = 1 - \frac{(0.4)}{1}$$

$$\text{Adjusted } R^2 = 0.6$$

Till Next Time...