

Course Summary

Course นี้ จะมีเป้าหมายในการสอน Machine Learning Model ที่สามารถนำไป Predict ข้อมูลที่เป็นตัวเลขได้ เช่น Predict ยอดขาย, Predict ราคาก่อน ซึ่งจะมีการสอน Regression Machine learning Model ที่หลากหลายรูปแบบ (มากกว่า 10 Model) โดยจะเริ่มด้วยการอธิบาย ทฤษฎีเบื้องต้นของ Model , Step การสร้าง Model , Hyperparameter ของ Model และทุก Model จะปิดท้ายโดยการเขียน R code ผ่าน Class R session เพื่อเป็นการสรุป ความรู้ของตัว Model นั้นๆ ซึ่งจำนวน Regression Machine Learning Model ทั้งหมดจะแบ่งออกเป็น 5 กลุ่มใหญ่ Linear Regression , Tree Regression , Support Vector Machine Regression , Distance Regression , Ensemble Regression. ซึ่งทุกคนที่มีความรู้เบื้องต้นของ สมการ คณิตศาสตร์ สามารถเรียน Course นี้ได้ Course ทั้งหมดจะประกอบไปด้วย 40 video ซึ่งมีความยาวรวมกันมากกว่า 7 ชั่วโมง

(โดยหากมีคำถามหรือข้อสงสัยสามารถถามผ่านผู้สอนมาได้ ซึ่งอาจจะนำคำถามเหล่านั้นมาสร้าง Video อธิบายเพิ่มเติมได้)

Course Objective

หลังจากเรียน Course นี้จบ นักเรียนสามารถ

- เข้าใจทฤษฎีของ regression machine learning model
- Step ของการสร้าง Regression Machine Model มากกว่า 10 model
- การเขียน R code ของ Regression Machine Model มากกว่า 10 model ซึ่งสามารถนำไปใช้ได้จริง
- พื้นฐานของ Machine Learning ทั้งการทำ Model และการทำ Preprocess ซึ่งสามารถนำไปใช้ได้จริงและไปต่อยอดได้ง่าย
- สามารถใช้ Regression เพื่อทำการทำนาย และการอธิบายผลของ Model ที่เราสร้างมาได้
- Course นี้เหมาะกับผู้ที่สนใจในสาย Data Science, Data Analytics หรือ สนใจที่จะนำ Machine Learning Model ไปใช้ในธุรกิจและชีวิตประจำวัน

Prerequisite

- คอมพิวเตอร์ส่วนตัว ที่สามารถลง R และ R studio ได้
- Basic สมการตัวเลข เบื้องต้น
- สามารถเขียน R เบื้องต้น หรือ เข้าใจแค่ Basic ของ R (หากยังไม่เคยมีประสบการณ์เลย สามารถเรียนฟรีที่ Course Intro to R ; DataRockie ได้)

<https://datarockie.teachable.com/p/r-programming>

Course Outline

11 Chapter (40 Episode, more than 7 hours)

1. General Introduction

1. Ep1 Getting Started
2. Ep2 Regression & ML Introduction

2. Simple Linear Regression

1. Ep3 Linear Regression Fundamental
2. Ep4 Term Definition in Linear Regression
3. Ep5 (Optional) Coefficient Prove for Simple Linear Regression
4. Ep6 (Optional) Coefficient Prove for Simple Linear Regression by Data

3. General Machine Learning Methodology

1. Ep7 Data Split
2. Ep8 Validation Set
3. Ep9 Hyperparameter Tuning
4. Ep10 Prediction & Evaluation
5. Ep11 Optimization

4. First R Session (Simple Linear Regression)

1. Ep12 R session Simple Linear Regression
2. Ep13 Interpreting Result Simple Linear Regression

5. Multiple Linear Regression

1. Ep14 Multiple Linear Regression
2. Ep15 Categorical Input
3. Ep16 (Optional) Coefficient Prove for Multiple Linear Regression
4. Ep17 Non-Linear Regression

6. Potential Problem of Complex Modeling

1. Ep18 Potential Problem of Complex Modeling
2. Ep19 Ridge Regression
3. Ep20 Lasso Regression
4. Ep21 Feature Scaling
5. Ep22 Elasticnet Regression
6. Ep23 R Session Multiple Linear Regression
7. Ep24 R Session Regularization Regression

7. Decision Tree Regression

1. Ep25 Various Type of Regression
2. Ep26 Introduction to Tree Regression
3. Ep27 Decision Tree Regression Steps
4. Ep28 Tree Pruning
5. Ep29 Random Forest Regression
6. Ep30 Gradient Boosting Tree Regression
7. Ep31 R Session Regression Tree

8. Support Vector Machine Regression

1. Ep32 Support Vector Machine Regression Linear Kernel
2. Ep33 Support Vector Machine Regression Kernel Trick
3. Ep34 R Session Support Vector Machine Regression

9. Distance Regression

1. Ep35 Distance Regression
2. Ep36 R Session Distance Regression

10. Ensemble Regression

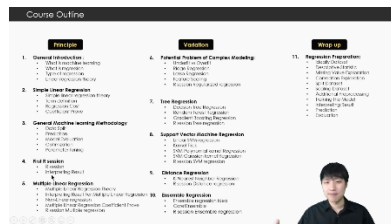
1. Ep37 Ensemble Regression
2. Ep38 R Session Ensemble Regression

11. Regression Preparation

1. Ep39 Wrapup Machine Learning Preprocess
2. Ep40 R Session Wrapup Machine Learning Process

Episode Description

Ep1 Getting Started:



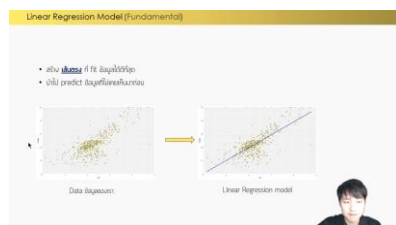
อธิบายถึงสำคัญของ Regression for machine learning , ความต้องการของ Data science , Machine learning skills ซึ่งจะมีการอธิบายโครงทั้ง Courses ในตอนนี้.

Ep2 Regression & ML Introduction



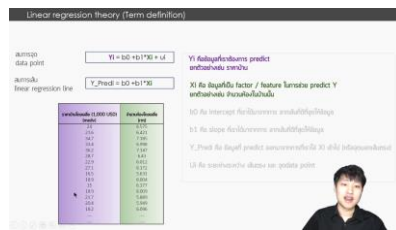
อธิบายถึง definition ต่างๆที่เราต้องเข้าใจ เช่น machine learning คืออะไร ,มีกี่ประเภท, regression คืออะไร รวมถึง use case ของ regression machine learning

Ep3 Linear Regression Fundamental



ในตอนนี้จะมีการอธิบายถึง พื้นฐานและ ทฤษฎีของ simple linear regression และการเขียน R code เบื้องต้นการ simple linear regression

Ep4 Term Definition in Linear Regression



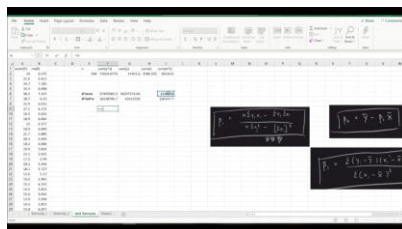
ความหมายของคำศัพท์ต่างๆ และความสำคัญของแต่ละตัว ที่จะต้องเจอในเนื้อหา simple linear regression เช่น Y_i , X_i , Cost, Betas, Etc. และ ความ related กับระหว่างคำศัพท์เหล่านั้นกับ ตัว model simple linear regression

Ep5 (Optional) Coefficient Prove for Simple Linear Regression

$$\beta_1 = \frac{n \sum y_i x_i - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}$$
$$\beta_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

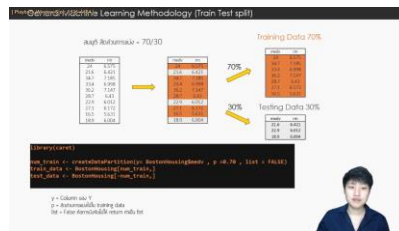
การแกะสูตร beta coefficient ของสมการ simple linear regression โดยใช้การ prove ทาง คณิตศาสตร์เป็นเนื้อหา optional สำหรับคนที่อยากเรียนรู้เรื่องการ prove beta และมีความรู้ด้าน math, stat พื้นฐาน เช่น การหา derivative

Ep6 (Optional) Coefficient Prove for Simple Linear Regression by Data



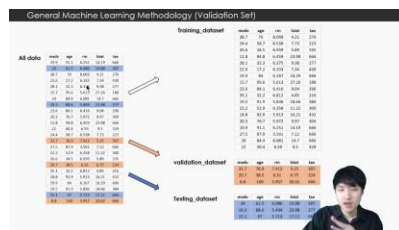
หลังจากเราได้สูตร beta coefficient ของ Ep5 มาแล้ว เราจะมาลองกับ data จริงว่าจะ return coefficient ให้ค่าเท่ากับใน R หรือไม่

Ep7 Data Split



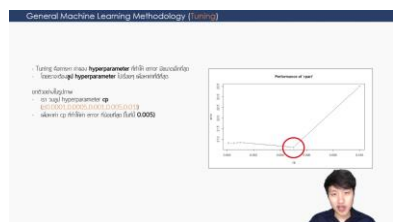
เป็น Topic หนึ่งใน general machine learning methodology (การทำ machine learning ต้องมี step พื้นฐานอย่างไรบ้าง) โดย data split จะเป็นการแบ่ง data เพื่อไป สร้าง model และ ประเมิน model

Ep8 Validation Set



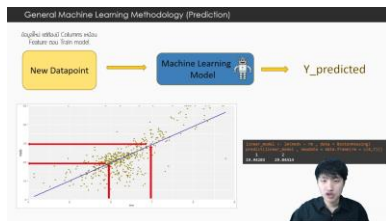
เป็น Topic หนึ่งใน general machine learning methodology โดย validation set จะเป็นการ add on ของ data split คือการแบ่ง data เพิ่มอีกหนึ่งก้อน

Ep9 Hyperparameter Tuning



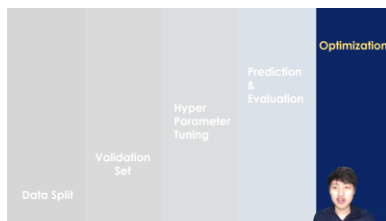
เป็น Topic หนึ่งใน general machine learning methodology โดย hyperparameter tuning จะเป็นวิธีในการหา best hyperparameter โดยการประเมินจาก error ของแต่ละ hyperparameter ที่ใส่เข้าไปใน model

Ep10 Prediction & Evaluation



เป็น Topic หนึ่งใน general machine learning methodology โดย Prediction & Evaluation จะเป็นการนำ model ที่ train มาไป ลอง test กับ ข้อมูลใหม่ แล้ววัดผลความสามารถของ model เรา

Ep11 Optimization



เป็น Topic หนึ่งใน general machine learning methodology โดย optimization จะเป็น วิธีการข้างหลัง ที่ model แต่ละตัวจะ return ผลลัพธ์ที่ดีที่สุดออกมา (Given Hyperparameter)

Ep12 R session Simple Linear Regression

```
## Simple Linear Regression
## Data
data <- read.csv("data.csv")
## Data Split
set.seed(123)
train_index <- sample(nrow(data), 0.7)
train_data <- data[train_index, ]
test_data <- data[-train_index, ]
## Train Model
model <- lm(y ~ x, data = train_data)
## Prediction
predicted <- predict(model, newdata = test_data)
```

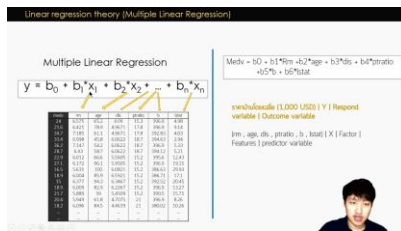
เราจะนำความรู้ที่เรียนไปก่อนหน้านี้มาใช้ใน module นี้โดยทำ Data split, Train model , Prediction and Evaluation ผ่านการเขียน R code

Ep13 Interpreting Result Simple Linear Regression



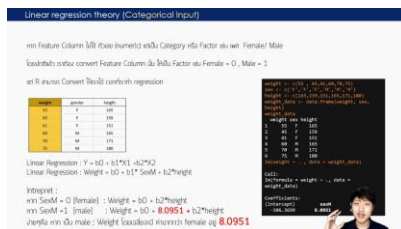
หลังจากเราทำ model simple linear regression in R ไปใน Ep12 (R session Simple Linear Regression) เราจะ
สามารถนำไป อธิบายผลเบื้องต้นได้ โดยใช้วิธีตาม Ep นี้

Ep14 Multiple Linear Regression



เราจะเริ่มเรียน multiple linear regression model ที่ซึ่งเป็น model ที่ complex มากขึ้น เมื่อเทียบกับ simple linear regression โดยจะอธิบายถึง พื้นฐาน , R Code และความแตกต่างระหว่าง 2 model นี้

Ep15 Categorical Input



บางที่ feature ที่เราใส่เข้าไปไม่ใช่ ตัวเลข (numerical) แต่เป็น หมวดหมู่ (categorical) แทน เช่น อากาศ ร้อน,เย็น เราจะมีวิธีอธิบายผลที่แตกต่างกันไป

Ep16 (Optional) Coefficient Prove for Multiple Linear Regression

The image shows a handwritten derivation for the coefficient β in multiple linear regression. It starts with the equation $u = (Y - X\beta)$ and then shows the derivative of the sum of squares with respect to β . The final result is $\beta = (X^T X)^{-1} X^T Y$.

ในตอนนี้จะมีการพูดถึง การแกะสูตร beta coefficient ของสมการ multiple linear regression โดยใช้การ prove ทางคณิตศาสตร์ เป็นเนื้อหา optional สำหรับคนที่อยากเรียนรู้เรื่องการ prove beta และมีความรู้ด้าน math , stat พื้นฐาน เช่น การหา derivative, linear algebra

Ep17 Non-Linear Regression

The slide is titled "Linear regression theory (Polynomial Regression)". It discusses the assumption that a linear regression model assumes a linear relationship between the response variable Y and the predictor X . It then introduces polynomial regression as a way to model non-linear relationships. The slide shows three equations for different polynomial degrees: $y = b_0 + b_1x$, $y = b_0 + b_1x + b_2x^2 + \dots + b_nx^n$, and $y = b_0 + b_1x + b_2x^2 + \dots + b_nx^n$. It also includes a graph showing a non-linear relationship and a small video thumbnail of a person.

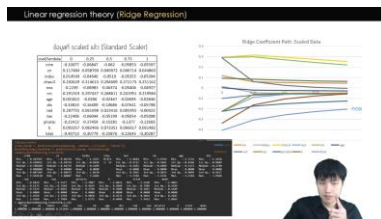
หาข้อมูลเราไม่มีความสัมพันธ์เชิงเส้นตรง เราจะสามารถใช้ model Nonlinear Regression ในการ Modeling ข้อมูลเราได้ ซึ่งใน class นี้จะเป็นตัว Polynomial Regression Model

Ep18 Potential Problem of Complex Modeling

The slide is titled "Potential Problem of Complex Modeling". It discusses the potential problems of complex modeling, including overfitting and underfitting. It shows three graphs: "Underfitting" (a straight line), "Overfitting" (a highly complex curve), and "Balance" (a smooth curve). It also includes a small video thumbnail of a person.

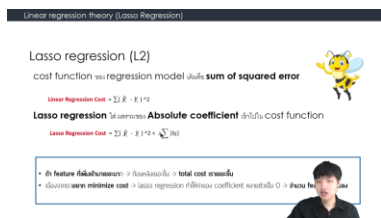
หากเราใส่ feature เข้ามาเยอะเกินจนทำให้ Model Complex อาจจะทำให้ Model เรามีปัญหาได้ ซึ่งประเภทของปัญหาที่พูดถึงจะถูกพูดใน Video นี้

Ep19 Ridge Regression



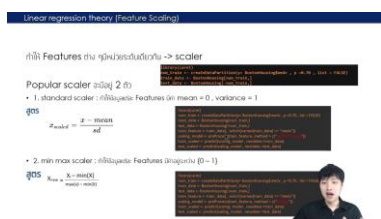
หนึ่งใน Machine Learning model ที่สามารถแก้ปัญหา Overfitting ได้ จะมีการอธิบายถึง ทฤษฎี , Hyperparameter, และ R code

Ep20 Lasso Regression



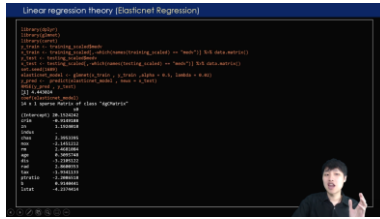
อีกหนึ่งใน Machine Learning Model ที่สามารถแก้ปัญหา Overfitting ได้ จะมีการอธิบายถึง ทฤษฎี , Hyperparameter, และ R code

Ep21 Feature Scaling



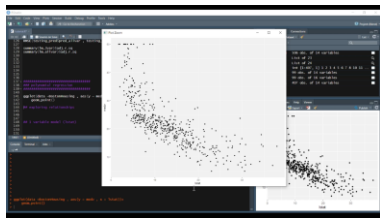
เป็นวิธีการ **Preprocess** หนึ่งที่สำคัญต่อการทำ **Regularization Machine Learning Model**, หรือแม้กระทั่ง **Machine Learning Model** ที่ ต้องการให้ หน่วยของข้อมูลอยู่ใกล้ๆ กัน

Ep22 Elasticnet Regression



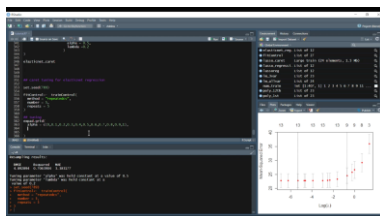
เป็นอีกหนึ่งใน Machine Learning Model ที่สามารถแก้ปัญหา Overfitting ได้ จะมีการอธิบายถึง ทฤษฎี , Hyperparameter, และ R code

Ep23 R Session Multiple Linear Regression



เราจะนำความรู้ที่เรียนไปในบทเรียน multiple linear regression และ nonlinear regression มาใช้ใน module นี้ทั้ง การ เขียน model แบบธรรมดา, แบบ caret, summary & prediction และ evaluation ผ่านการเขียน R code

Ep24 R Session Regularization Regression



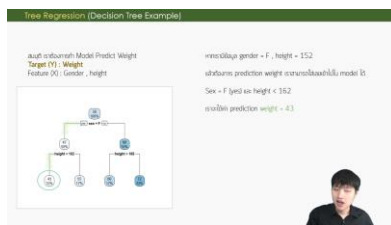
เราจะนำความรู้ที่เรียนไปในบทเรียน Ridge regression, Lasso regression , Elasticnet regression, Feature scaling มาใช้ใน Module นี้ทั้งการ เขียน model ทั้ง 3 แบบผ่าน package glmnet, caret, การ Tune model ให้ได้ Model ที่ดี ที่สุด และการ Evaluation

Ep25 Various Type of Regression



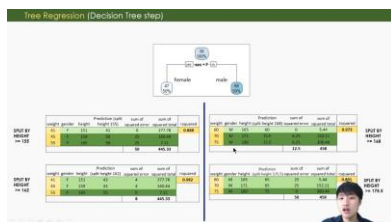
เราได้เรียนถึง Regression Machine Learning ประเภท Linear regression แต่ที่จริงแล้วยังมี Machine Learning Model หลายตัวที่สามารถทำ Regression Problem ได้เหมือนกัน เราจะมาพูดถึง Machine Learning Model ทั้งหมดที่จะสอนกันใน course นี้

Ep26 Introduction to Tree Regression



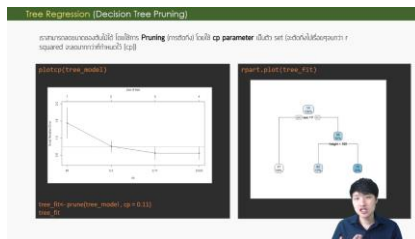
class นี้เราจะพูดถึง model ตัวใหม่ที่เป็น tree-based model ซึ่งจะมีหลักการคิดไม่เหมือนกับเจ้าตัว linear regression เลย และจะพูดถึง term definition ต่างๆ ที่จะใช้ใน tree based model

Ep27 Decision Tree Regression Steps



class นี้เราจะพูดถึง การสร้าง decision tree regression step by step เพื่อเพิ่มความเข้าใจ model นี้แบบ ละเอียดมากขึ้น

Ep28 Tree Pruning



หลังจากที่เราได้สร้าง model decision tree regression แล้ว เราจะสามารถทำ pruning เพื่อแก้ปัญหาด้าน overfit ได้

Ep29 Random Forest Regression



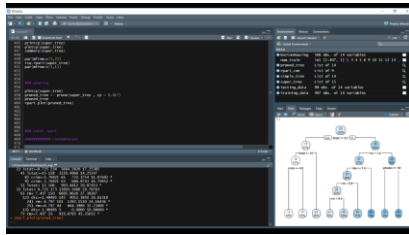
เป็น machine learning model ตัวใหม่ที่เป็น tree based แต่เป็น model ที่ฮิตมากและค่อนข้างใช้งานง่าย โดยจะมีการอธิบายถึง ข้อแตกต่างกับ regression tree ปกติ, ขั้นตอนการสร้าง random forest model และ การ อธิบายผลลัพธ์ที่ได้จาก model

Ep30 Gradient Boosting Tree Regression



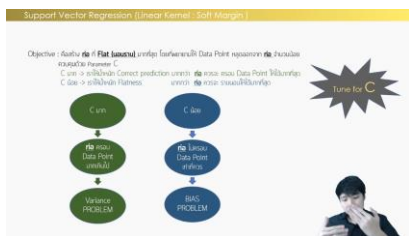
เป็นอีกหนึ่ง model ใหม่ที่เป็น tree based แต่มีความสามารถที่จะให้ ผลprediction ที่แม่นยำ โดยจะมีการอธิบายถึง ข้อแตกต่างกับ random forest model , step การสร้าง gbm แบบละเอียด และ การ อธิบายผลลัพธ์ที่ได้จาก model

Ep31 R Session Regression Tree



เราจะนำความรู้ที่เรียนไปในบทเรียน decision tree regression, random forest regression , gradient boosting machine regression มาใช้ใน module นี้ ทั้งการ เขียนmodel ทั้ง 3 แบบผ่าน package rpart, ranger, gbm , caret การ pruning regression tree, การ tune model ให้ได้ model ที่ดีที่สุด และการ evaluation

Ep32 Support Vector Machine Regression Linear Kernel



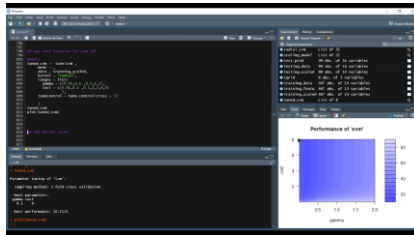
class นี้เราจะพูดถึง model ตัวใหม่ที่เป็น Support Vector machine based model ซึ่งจะมีหลักการคิดไม่เหมือนกับเจ้าตัว linear regression และ decision tree เลย และจะพูดถึง ทฤษฎี ต่างๆ, hyperparameter ที่จะใช้ใน support vector machine linear kernel และการเขียน R code

Ep33 Support Vector Machine Regression Kernel Trick

y	x	x^2	x^3	x^4
12	1	1	1	1
8	1.5	2.25	3.375	5.0625
6	2	4	8	16
4	3	9	27	81
3	4	16	64	256
2	6	36	216	1296
1.5	8	64	512	4096
1	12	144	1728	20736

จะเป็น model ใหม่ที่เป็น support vector machine based 2 ตัว , Polynomial kernel และ RBF kernel ซึ่งเป็น model ที่มีความสามารถที่จะให้ค่า prediction ที่แม่นยำขึ้น จะมีการพูดถึง step การสร้าง Polynomial kernel , Hyperparameter ของทั้ง 2 model และ การเขียน R code

Ep34 R Session: Support Vector Machine Regression



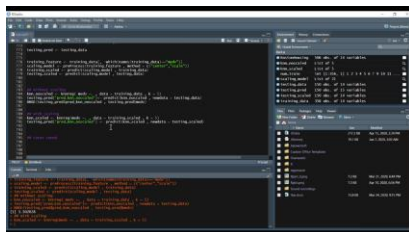
เราจะนำความรู้ที่เรียนไปใน Class Support Vector Machine Regression Linear Kernel , Kernel trick มาใช้ใน Module นี้ทั้งการเขียน Model ทั้ง 3 แบบผ่าน package e1071, kernlab , caret การ Tune model ให้ได้ model ที่ดีที่สุดและการ Evaluation

Ep35 Distance Regression



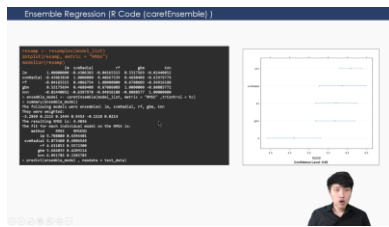
Class นี้เราจะพูดถึง Model ตัวใหม่ที่เป็น distance based: K-nearest neighbor regression ซึ่งจะมีหลักการคิดไม่เหมือนกับ Model Linear regression, Decision Tree Regression, SVM regression. อธิบายถึงทฤษฎี ต่างๆ, Hyperparameter ที่จะใช้ใน KNN Regression และการเขียน R code

Ep36 R Session Distance Regression



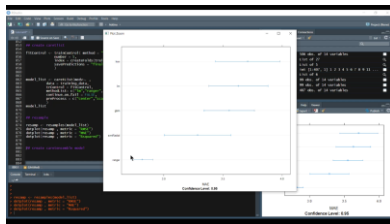
เราจะนำความรู้ที่เรียนไปใน class Distance Regression มาใช้ใน Module นี้ ทั้งการเขียน KNN model การ Tune Model ให้ได้ Model ที่ดีที่สุด และการ Evaluation

Ep37 Ensemble Regression



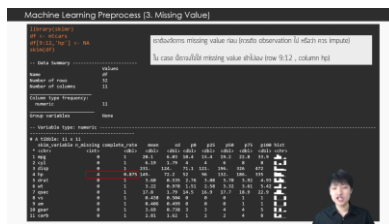
Model ครั้งนี้จะเป็นการรวบรวมพวก Prediction จาก Model ต่างๆ หลาย Model เพื่อหาค่า Prediction Out of Sample ที่ดีที่สุดโดยใช้ Model Ensemble ซ้อนทับผล Prediction จาก Model Tree Regression, Linear Regression, SVM Regression , KNN regression เพื่อรวบรวมผลให้เป็น Final Prediction Model

Ep38 R Session Ensemble Regression



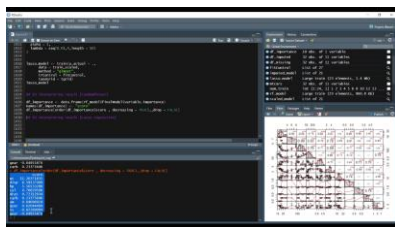
เราจะนำความรู้ที่เรียนไปใน Class Ensemble Regression มาใช้ใน module นี้ ทั้งการ เขียน Ensemble Model ผ่าน `caretEnsemble` Package , การ Explore ผลลัพธ์ใน Ensemble Model และการ Evaluation

Ep39 Wrapup Machine Learning Preprocess



Class นี้จะรวบรวม Step เบื้องต้นก่อนที่เราจะทำ Regression Machine Learning Model เช่น การ Explore Data, Explore Missing value, Explore Correlation, Scaling Data รวมไปถึง Step การทำ Regression Machine Learning model เบื้องต้น และ Post Model Analysis เพื่อที่จะเป็น Guideline ในการ Applied Regression Machine Learning model กับ dataset ใน ชีวิตจริง

Ep40 R Session Wrapup Machine Learning Process



เราจะนำความรู้ที่เรียนไปใน class wrapup Machine Learning Preprocess มาใช้ใน class นี้ ตั้งแต่ Step หลังจากการเรียก Data ไปจนถึงการ Evaluation เพื่อเป็น guideline code ในประยุกต์ Regression For Machine Learning ในชีวิตจริงต่อไป