GENERAL ASSEMBLY

DS-DC-13 LECTURE NOTES

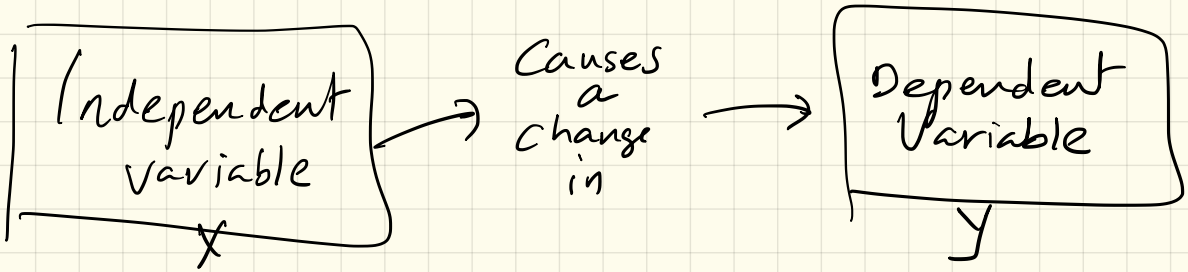
Jim Simpson
Data Scientist, Sotera

LESSON 06

LESSON 6:
LINEAR REGRESSION

LEARNING OBJECTIVES

- [ ] DEFINE SIMPLE LINEAR REGRESSION
- [ ] BUILD A LINEAR REGRESSION MODEL USING SCIKIT-LEARN
- [ ] UNDERSTANDING MULTICOLLINEARITY IN A MULTIPLE REGRESSION

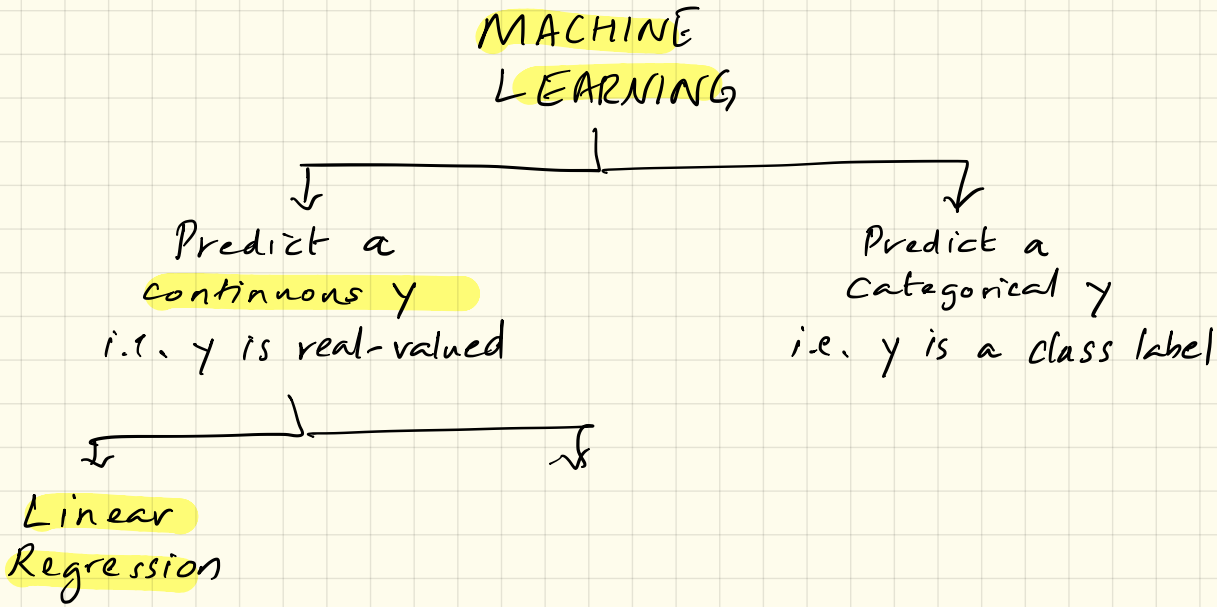


X

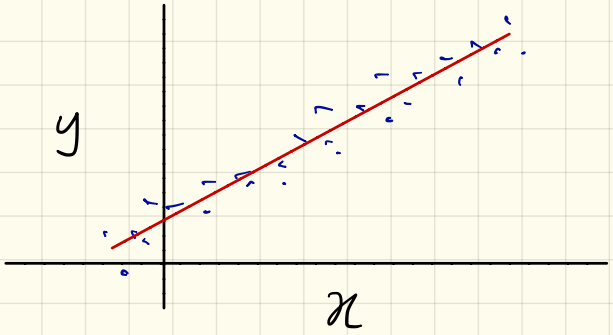
Input
Feature
Factor (categorical)
Covariate (continuous)
Cause
Predictor
Explanatory
Manipulated
Controlled

Y

Output
Prediction
Effect
Outcome
Explained
Measured
Responding


MACHINE LEARNING
Predict a continuous y
i.e. y is real-valued
Predict a categorical y
i.e. y is a class label
Linear Regression

SIMPLE LINEAR REGRESSION



$y = mx + c$

m: slope, c: intercept

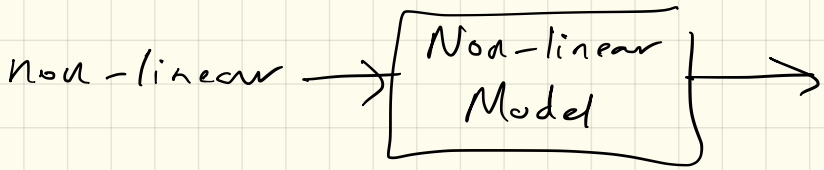
$y = \beta X + \alpha$

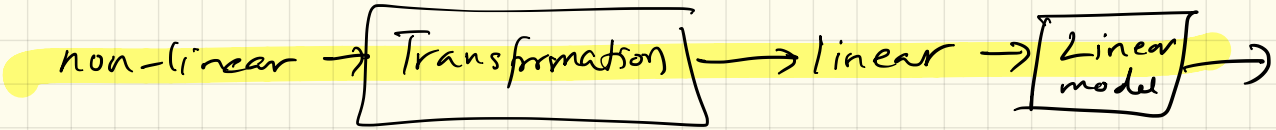
β: coefficients

	x	y
1	.	.
2	.	.
3	.	.

	X			
	x_1	x_2	x_3	y
1				
2				
3				

Addressing non-linearity





Normalization

Standardization

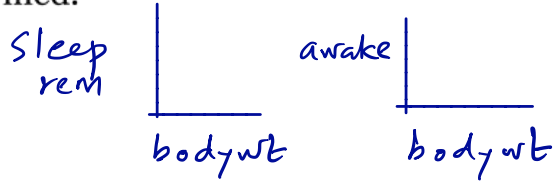
$\mu = 0 \quad \sigma = 1$

ACTIVITY: GENERATE SINGLE VARIABLE LINEAR MODEL PLOTS

EXERCISE

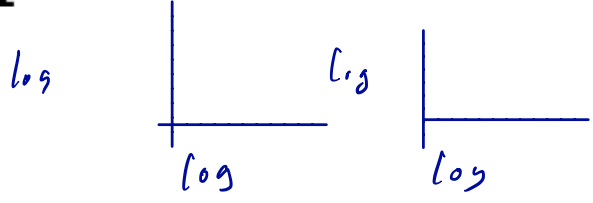
DIRECTIONS (15 minutes)

1. Update and complete the code in the starter notebook to use **lmplot** and display correlations between body weight, **bodywt** and two dependent variables: **sleep_rem** and **awake**.
2. For each, generate linear models for the variables as-is and log-transformed.

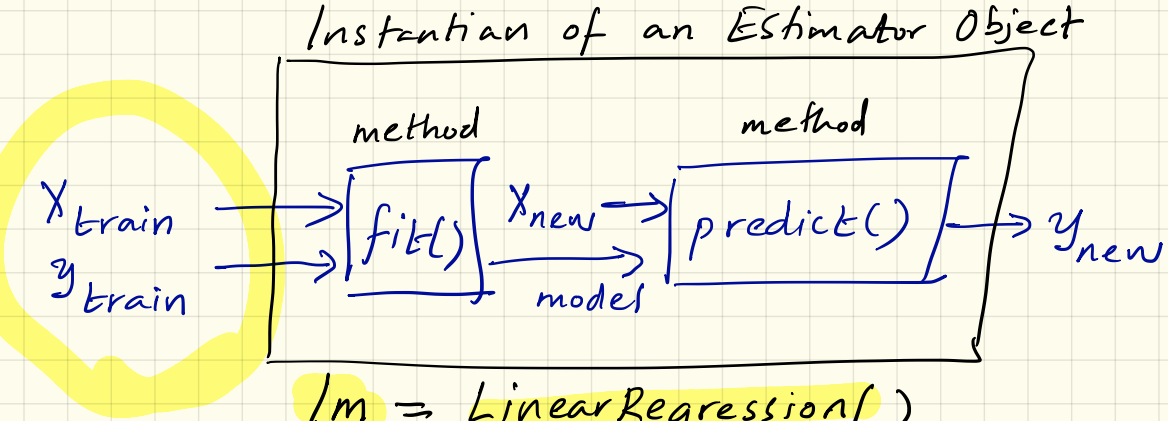


DELIVERABLE

Two plots



SCIKIT-LEARN FIT-PREDICT MODEL

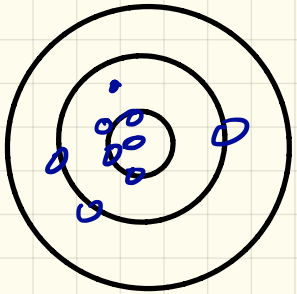


lm = LinearRegression()

Instantiation

Object


-1.5
0
2.5



vector	x	$y = mx + c$
matrix	X	$y = \beta X + \alpha$