

Spatial Data Science

Machine Learning for Everyone

(EPA122A)

Lecture 5

Theodoros Chatzivasileiadis

THIS IS YOUR MACHINE LEARNING SYSTEM?

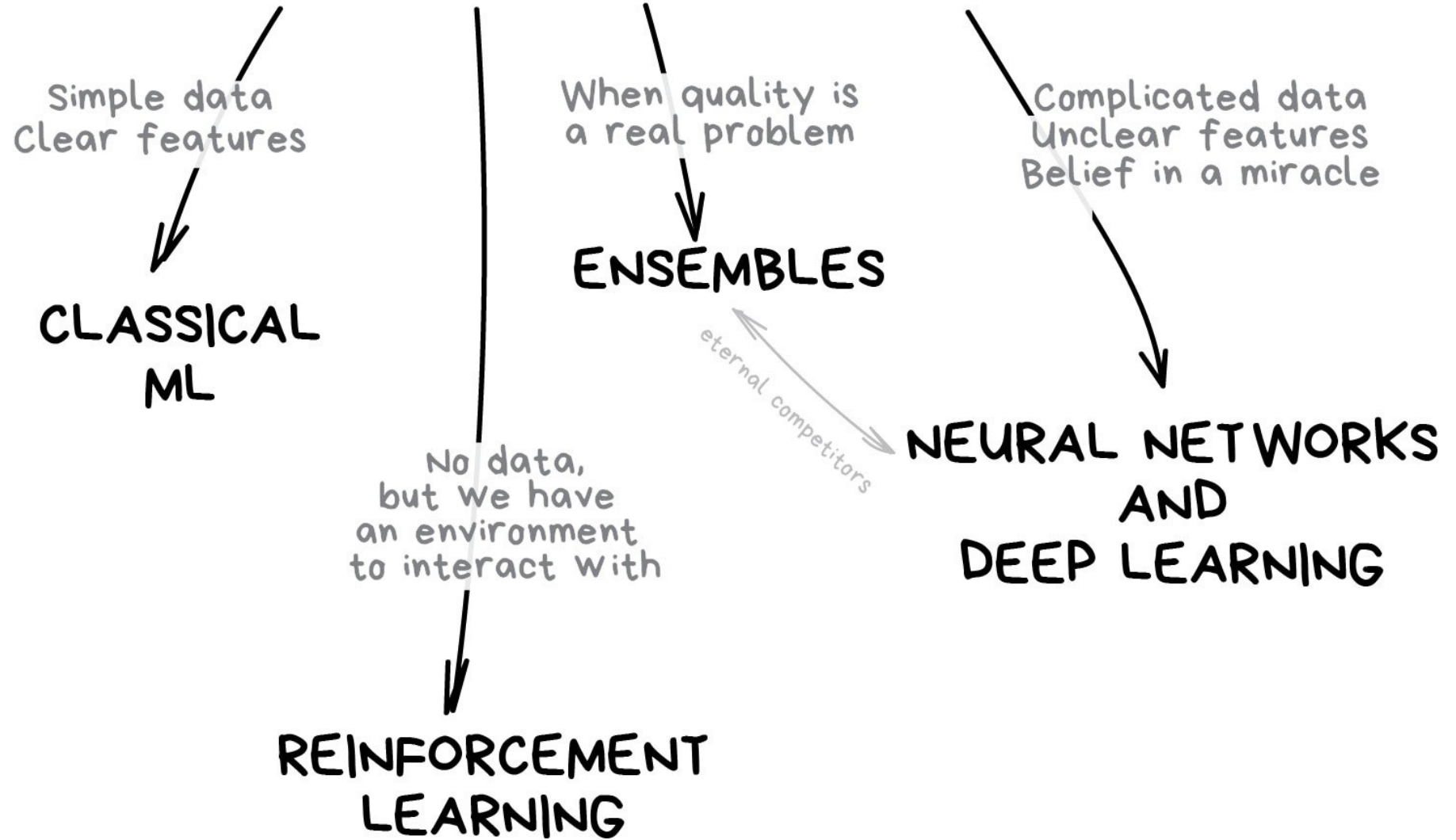
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

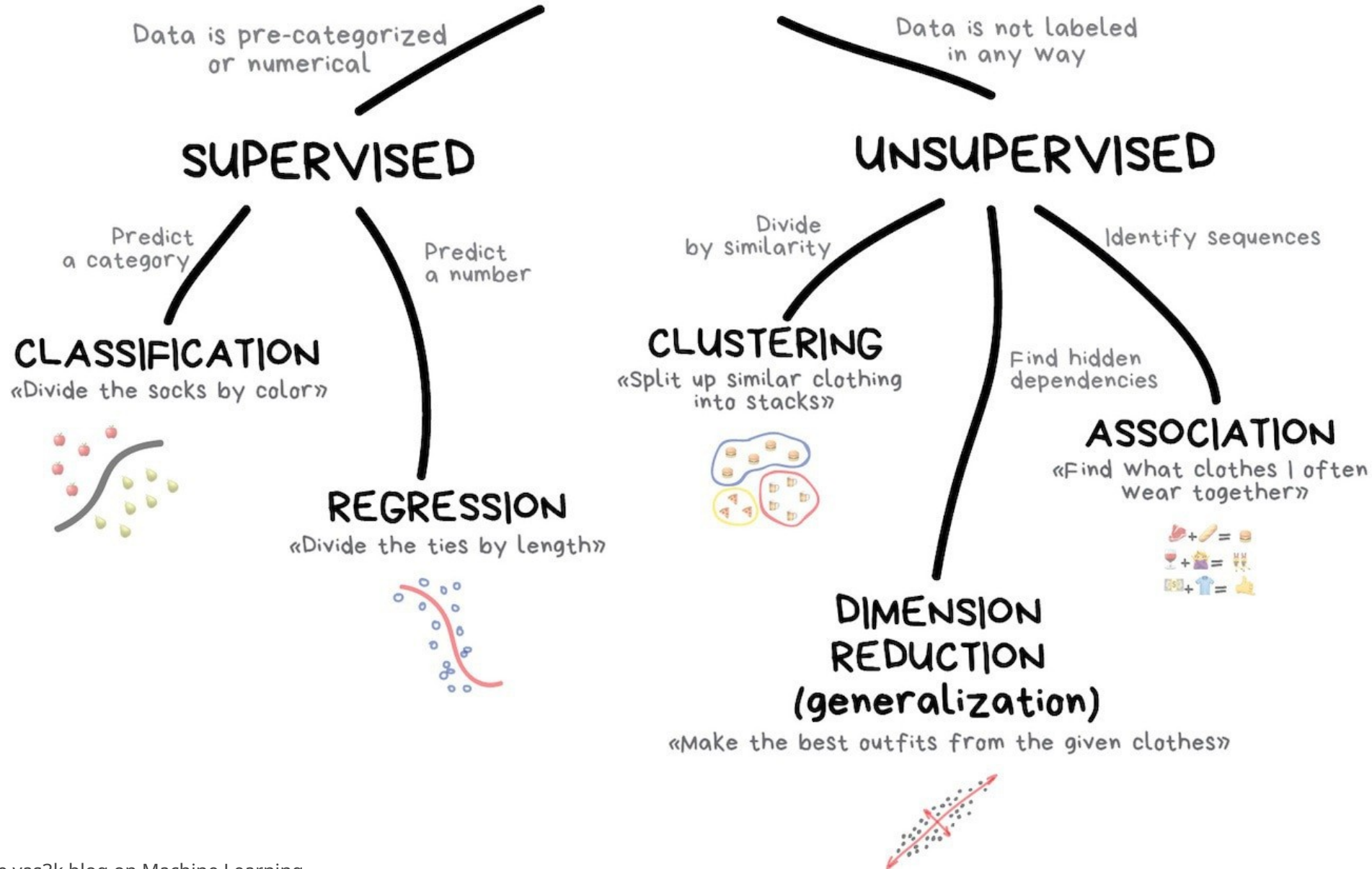
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



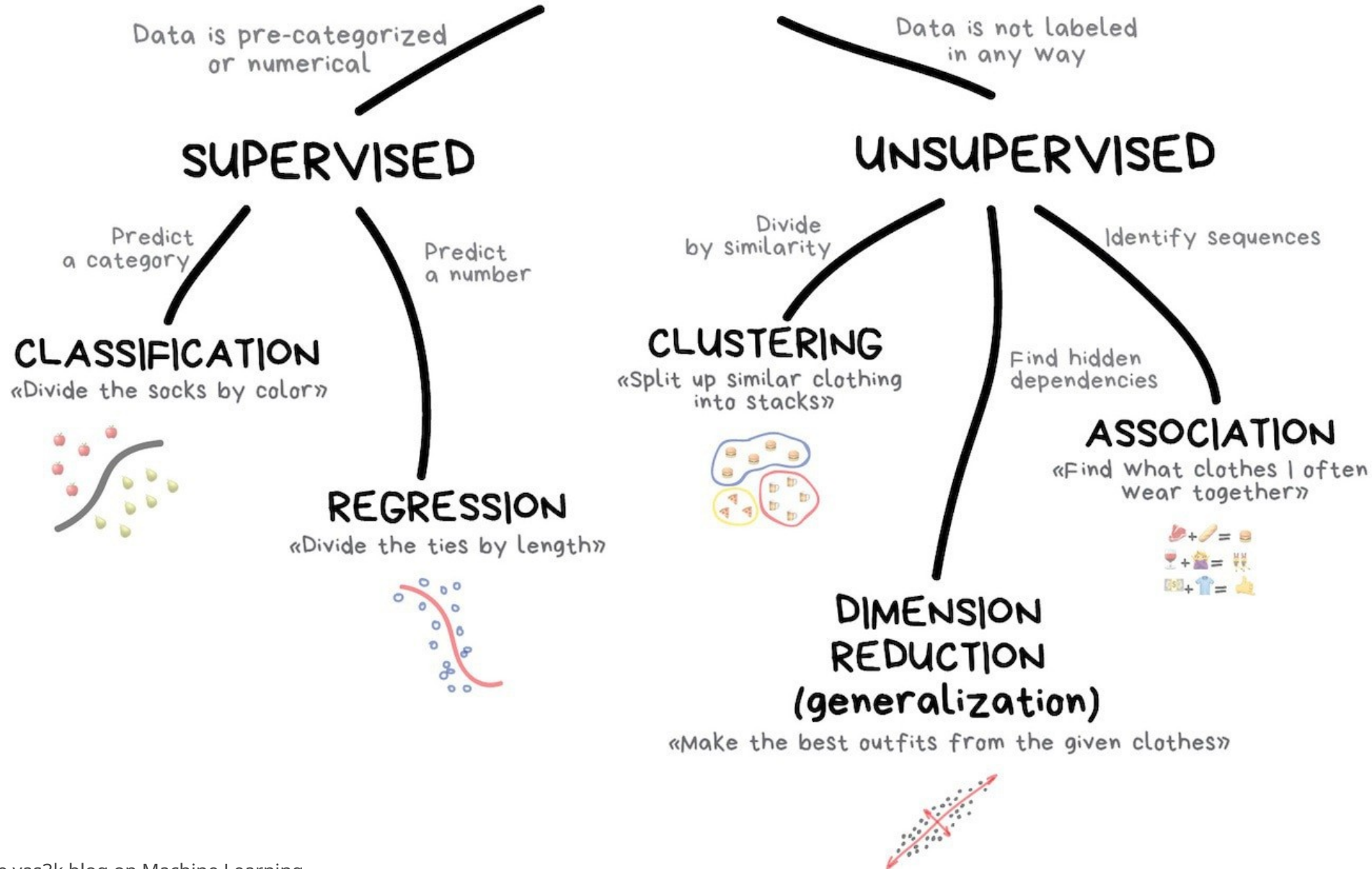
THE MAIN TYPES OF MACHINE LEARNING



CLASSICAL MACHINE LEARNING



CLASSICAL MACHINE LEARNING



What is a regression model

Regression analysis is a statistical method that helps us understand and predict how different factors are related. By analyzing data, it shows how changes in one variable (like population density) might affect another (such as traffic flow or housing demand).

What is a regression model

Regression analysis is a **statistical method** that helps us understand and predict how different factors are related. By analyzing data, it shows how changes in one variable (like population density) might affect another (such as traffic flow or housing demand).

What is a regression model

Regression analysis is a statistical method that helps us **understand and predict** how different factors are related. By analyzing data, it shows how changes in one variable (like population density) might affect another (such as traffic flow or housing demand).

What is a regression model

Regression analysis is a statistical method that helps us understand and predict **how different factors are related**.

What is a regression model

Regression analysis is a statistical method that helps us understand and predict **how different factors are related**.

By analyzing data, it shows how changes in one variable (like population density) might affect another (such as traffic flow or housing demand).

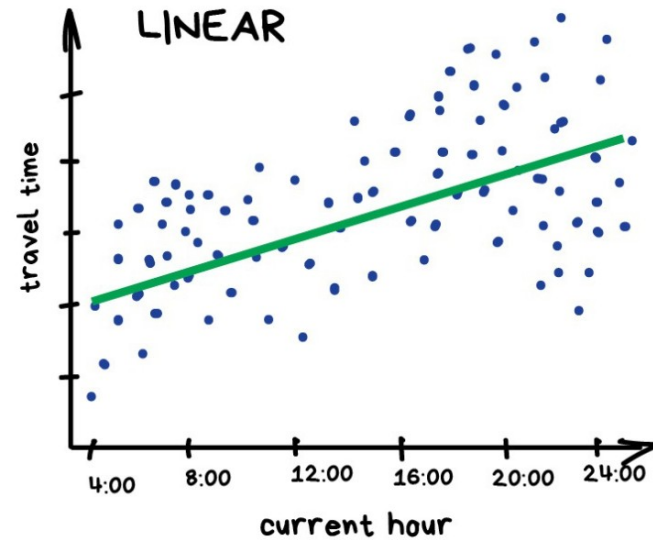
This tool enables us to make **informed** decisions by forecasting outcomes and identifying key influences in urban design and planning.

- Regression analysis is a statistical method that helps us understand and predict **how different factors are related**.
- Each regression model has **at least** two variables:
 1. A Dependent Variable: Always referred as y or Y
 2. An Independent Variable: Referred as x.

- **Dependent Variable:** This is the outcome you're interested in predicting or understanding. It's called "dependent" because its value depends on other factors. For architects and spatial planners, this could be variables like property values, energy consumption of a building, or the level of foot traffic in a public space.
- **Independent Variables:** These are the factors that you suspect have an influence on the dependent variable. They are "independent" because they are presumed to cause or explain changes in the dependent variable, **not the other way around**. Examples include the distance to public transportation, availability of green spaces, building materials used, or population density.

In essence, the regression model helps you analyze how changes in independent variables affect the dependent variable, enabling you to make data-driven decisions in your projects.

PREDICT TRAFFIC JAMS



$$Y_i = \alpha_0 + \beta_1 * x_i + \varepsilon_i$$

When the line is straight — it's a linear regression, when it's curved – polynomial. If the model has more than 1 dependent variables it is a multivariate regression model

$$Y_i = \alpha_0 + \beta_1 * x_{1i} + \varepsilon_i$$

- Y_i : The dependent Variable (Travel time)
- α_0 : The intercept, or constant
- β_1 : The Coefficient of x on Y
- x_{1i} : The independent variable (Time of the day)
- ε_i : The error term or residual

Response vs. Predictor Variables

$X = X_1, \dots, X_p$
 $X_j = x_{1j}, \dots, x_{ij}, \dots, x_{nj}$
Independent

$Y = y_1, \dots, y_n$
outcome
response variable
dependent variable

i observations

TV budget	Radio budget	Newspaper budget	sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9

p predictors

$$Y_i = \alpha_0 + \beta_1 * x_{1i} + \varepsilon_i$$

TV budget	sales
230.1	22.1
44.5	10.4
17.2	9.3
151.5	18.5
180.8	12.9

- Here what we are asking is: How much does TV budget (if any) does effect sales of a product.

$$Sales_i = \alpha_0 + \beta_1 * TV_Budget_i + \varepsilon_i$$

IF Sales = 0 if TV_Budget =0, the $\alpha_0 = 0$. In our case however sales would not be 0 without advertisement, so we have an intercept. That is the number of sales **irrespective** of the independent variable.

OLS Regression Results

```

=====
Dep. Variable:          y      R-squared:          0.750
Model:                  OLS    Adj. R-squared:      0.500
Method:                 Least Squares    F-statistic:      3.000
Date:                   Mon, 29 Aug 2022    Prob (F-statistic): 0.333
Time:                   16:37:40    Log-Likelihood:    -2.0007
No. Observations:       3      AIC:              8.001
Df Residuals:           1      BIC:              6.199
Df Model:                1
Covariance Type:        nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	0.6667	1.247	0.535	0.007	-15.181	16.514
x1	1.0000	0.577	1.732	0.333	-6.336	8.336
x2	-754.0001	676.067	XXXX	0.090	-800	-659

```

=====
Omnibus:               nan    Durbin-Watson:      3.000
Prob(Omnibus):         nan    Jarque-Bera (JB):    0.531
Skew:                  -0.707  Prob(JB):            0.767
Kurtosis:              1.500  Cond. No.             6.79
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.