LINEAR REGRESSION

Machine learning

Overview

WHAT IS LINEAR REGRESSION?

A linear regression is a data plot that graphs the linear relationship between an independent and a dependent variable(s).

It is typically used to visually show the strength of the relationship, and the dispersion of results.

E.g. to see test the strength of the relationship between amount of ice cream eaten and obesity.

Take the independent variable, the amount of ice cream, and relate it to the dependent variable, obesity, to see if there was a relationship.

MATHEMATICAL FORM

Linear Regression: Single Variable

$$\hat{y} = \beta_0 + \beta_1 x + \epsilon$$

Predicted output — Coefficients — Input — Error

Linear Regression: Multiple Variables

$$\hat{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

WHAT DO WE USE LINEAR REGRESSION FOR?

The overall idea of linear regression is to examine 2 things:

- Does a set of predictor variables do a good job in predicting an outcome (dependent) variable?
- Which variables in particular are significant predictors of the outcome variable and in what way do they *–indicated by the magnitude and sign of the beta estimates–* impact the outcome variable?

KEY POINTS …

- When selecting the model for the analysis, an important consideration is model fitting.
- Adding independent variables to a linear regression model will always increase the explained variance of the model (typically expressed as R^2).
- overfitting can occur by adding too many variables to the model, which reduces model generalizability.
- A simple model is usually preferable to a more complex model.
- Statistically, if a model includes a large number of variables, some of the variables will be statistically significant due to chance alone.

SIMPLE LINEAR REGRESSION

A college bookstore must order books 2 months before each semester starts. They believe that the number of books that will be sold for any particular course is related to the number of students registered for the course when the books are ordered.

They would like to develop a linear regression equation to help plan how many books to order.

From past records, the bookstore obtains the number of students registered, X, and the number of books actually sold for a course, y for 12 different semesters.

Semester	No of students	Books
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

WHAT IS THE ERROR TERM?

- An error term is a variable in a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables.

- The error term is also known as the residual, disturbance, or remainder term.

WHAT IS THE ERROR TERM?

- Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed.

- In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.

- *Points that do not fall directly on the trend line exhibit the fact that the dependent variable, in this case, the price, is influenced by more than just the independent variable, representing the passage of time.*

- *The error term stands for any influence being exerted on the price variable, such as changes in market sentiment.*

ERROR CALCULATION

The residual is calculated as: $r_i = y_i - \hat{y}$

where
r_i = residual value
y_i = observed value for a given x value
$\hat{y}$ = predicted value for a given x value

- The magnitude of a typical residual can give us a sense of generally how close our estimates are.
- The smaller the residual standard deviation, the closer is the fit to the data.
- In effect, the smaller the residual standard deviation is compared to the sample standard deviation, the more predictive, or adequate, the model is.

linear equation is $\hat{y} = 1x + 2$, the residual for each observation can be found.

For the first set, the actual y value is 1, but the predicted y value given by the equation is $\hat{y} = 1(1) + 2 = 3$. The residual value is, therefore, $1 - 3 = -2$, a negative residual value

X	Y	$\hat{y}$	error	$error^2$
1	1	3	-2	4
2	4	4	0	0
3	6	5	1	1
4	7	6	1	1

Sum of squared residuals: 6
Number of residuals less 1: 4 – 1 = 3
Residual standard deviation: $\sqrt{(6/3)} = \sqrt{2} \approx 1.4142$

LOSS FUNCTION

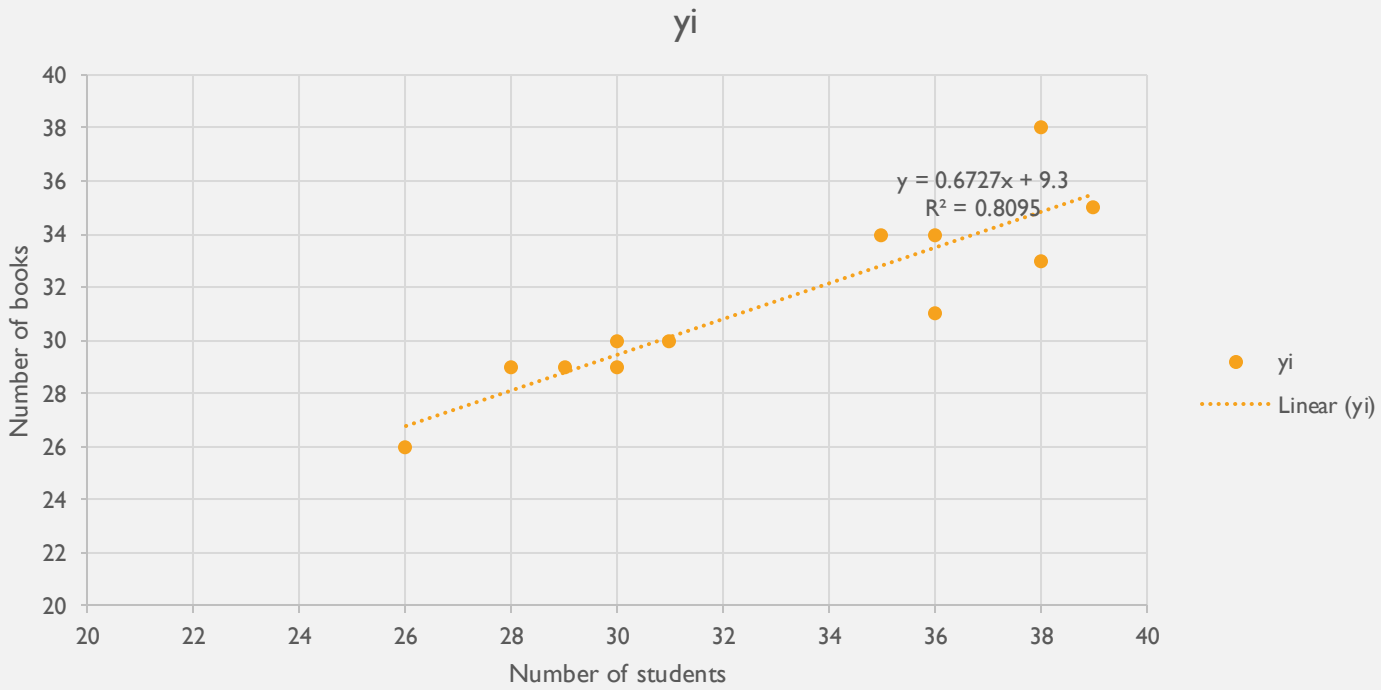
- Every machine learning problem is basically an optimization problem.
- The function that you want to optimize is usually called the loss function (or cost function).
- For the house price prediction example, after the model is trained, we are able to predict new house prices based on their features.
- For each house price we predict, denoted as $\hat{Y}i$, and the actual house price Yi we can calculate the loss by: $loss_i = (\hat{Y}_i - Y_i)^2$

$$L = \sum(\hat{Y}_i - Y_i)^2$$

- Which simply defines that our model's loss is the sum of distances between the house price we've predicted and the ground truth.
- This loss function, in particular, is called quadratic loss or least squares.
- We wish to minimize the loss function (L) as much as possible so the prediction will be as close as possible to the ground truth.

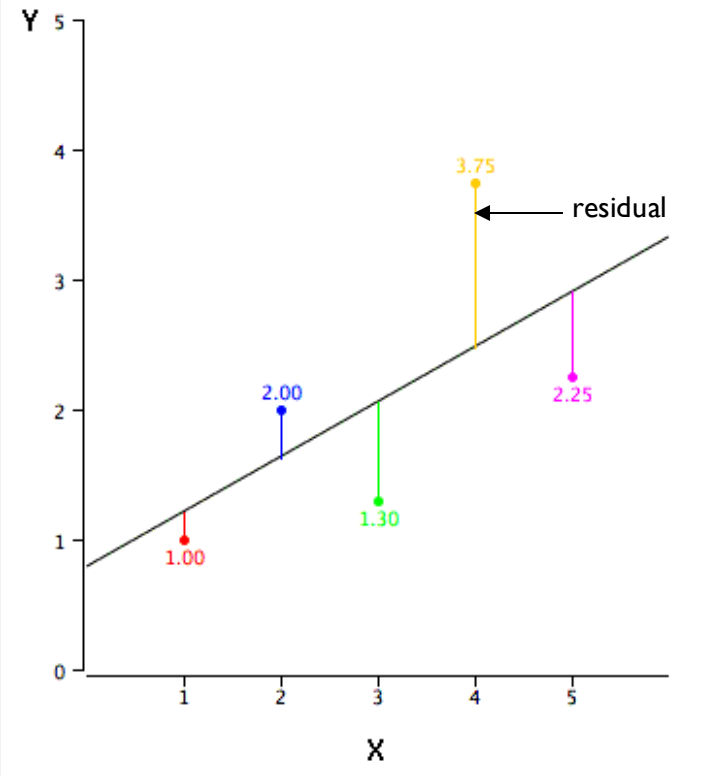
REGRESSION LINE

Obtain a scatter plot of the number of books sold versus the number of registered students.



COMPARING MODEL PREDICTIONS AGAINST REALITY

- Response = (Constant + Predictors) + Error
- Response/output is estimated for any given input or set of inputs
- Check these estimated outputs against the actual values
- The difference between the actual value and the model's estimate a residual.
- These residuals will play a significant role in judging the usefulness of a model.
- If the residuals are small, it implies that the model is a good estimator



MEAN ABSOLUTE ERROR

- Calculate the residual for every data point, taking only the absolute value of each so that negative and positive residuals do not cancel out.
- Take the average of all these residuals.
- Effectively, MAE describes the typical magnitude of the residuals.

$$MAE = \frac{1}{n}\sum\left|y - \hat{y}\right|$$

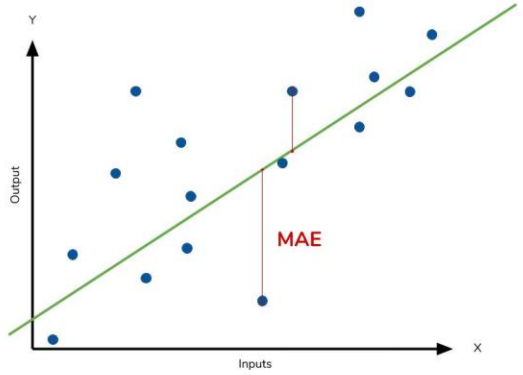
Divide by the total number of data points

Predicted output value

Actual output value

Sum of

The absolute value of the residual



MEAN ABSOLUTE ERROR

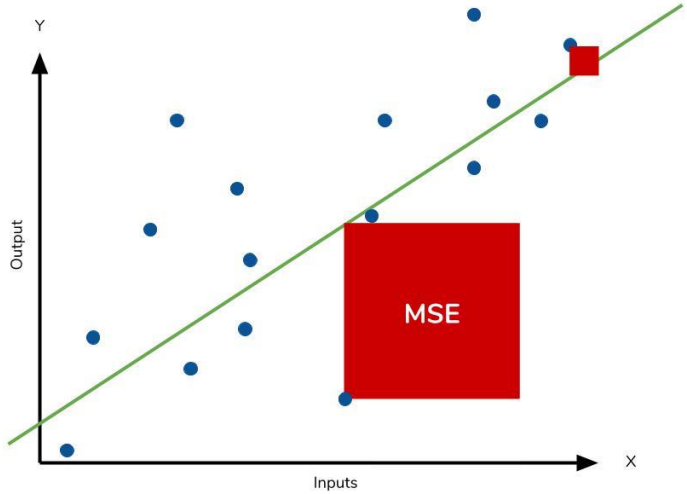
- The MAE is also the most intuitive of the metrics
- Each residual contributes proportionally to the total amount of error, meaning that larger errors will contribute linearly to the overall error.
- A small MAE suggests the model is great at prediction, while a large MAE suggests that your model may have trouble in certain areas.
- A MAE of 0 means that your model is a perfect predictor of the outputs (but this will almost never happen).

MEAN SQUARE ERROR

- The mean square error (MSE) is just like the MAE but squares the difference before summing them all instead of using the absolute value.

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \hat{y}\right)}^{2}$$

The square of the difference between actual and predicted



MEAN SQUARE ERROR

- Because we are squaring the difference, the MSE will almost always be bigger than the MAE.
  - For this reason, we cannot directly compare the MAE to the MSE.

- The effect of the square term in the MSE equation is most apparent with the presence of outliers in our data.

- Each residual in MAE contributes proportionally to the total error, the error grows quadratically in MSE.

- Outliers in our data will contribute to much higher total error in the MSE than they would the MAE.

ROOT MEAN SQUARED ERROR (RMSE)

- Square root of the MSE
- Since the MSE and RMSE both square the residual, they are similarly affected by outliers.
- The RMSE is analogous to the standard deviation (MSE to variance)
- Is a measure of how large your residuals are spread out.

$$\mathrm{MSE}(a) > \mathrm{MSE}(b) \iff \mathrm{RMSE}(a) > \mathrm{RMSE}(b)$$

MEAN ABSOLUTE PERCENTAGE ERROR

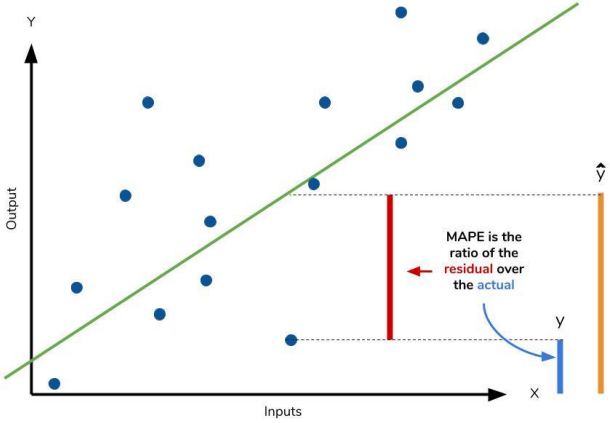
- The mean absolute percentage error (MAPE) is the percentage equivalent of MAE.
- The equation looks just like that of MAE, but with adjustments to convert everything into percentages.
- Both MAPE and MAE are robust to the effects of outliers thanks to the use of absolute value.

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

Multiplying by 100% converts to percentage

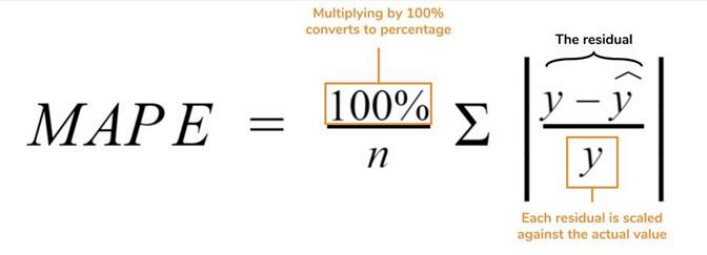
The residual

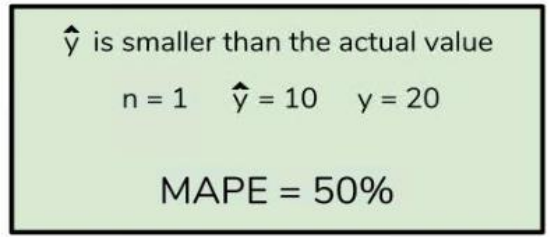
Each residual is scaled against the actual value

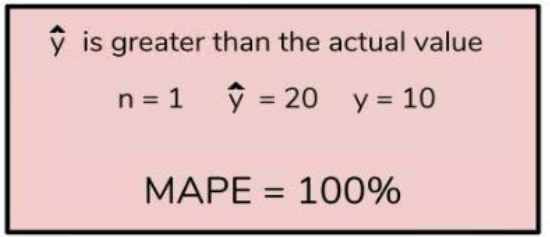


MEAN ABSOLUTE PERCENTAGE ERROR

- MAPE's weaknesses actually stem from use of division operation.
- MAPE is undefined for data points where the value is 0.
- Similarly, the MAPE can grow unexpectedly large if the actual values are exceptionally small themselves.
- Finally, the MAPE is biased towards predictions that are systematically less than the actual values themselves. That is to say, MAPE will be lower when the prediction is lower than the actual compared to a prediction that is higher by the same amount.







MEAN PERCENTAGE ERROR

- The mean percentage error (MPE) equation is exactly like that of MAPE.
- The only difference is that it lacks the absolute value operation.

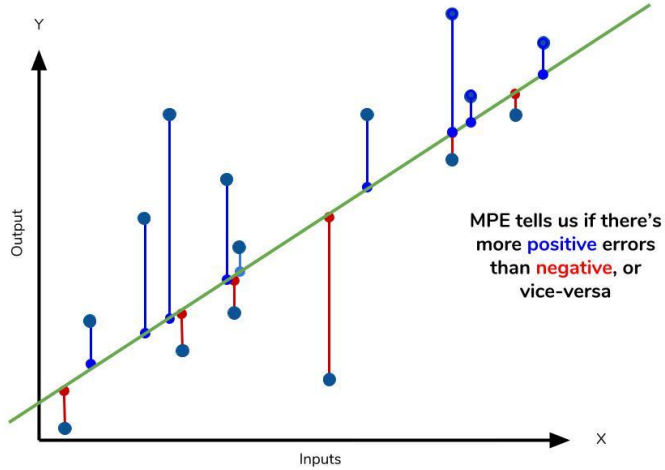
$$MPE = \frac{100\%}{n} \Sigma \left(\frac{y - \hat{y}}{y} \right)$$

$$MAPE = \frac{100\%}{n} \Sigma \left| \frac{y - \hat{y}}{y} \right|$$

Multiplying by 100% converts to percentage

The residual

Each residual is scaled against the actual value



MEAN PERCENTAGE ERROR

- Since positive and negative errors will cancel out, we cannot make any statements about how well the model predictions perform overall.
- However, if there are more negative or positive errors, this bias will show up in the MPE.
- Unlike MAE and MAPE, MPE is useful to us because it allows us to see if our model systematically underestimates (more negative error) or overestimates (positive error).

RESIDUAL PLOT

- After you fit a regression model, it is crucial to check the residual plots.
- If the plots display unwanted patterns, you can't trust the regression coefficients and other numeric results.
- Residual plots display the residual values on the y-axis and fitted values, or another variable, on the x-axis.

ESSENTIAL PARTS OF A REGRESSION MODEL

Dependent Variable = (Constant +Independent Variables) + Error

Or:

Dependent Variable = Deterministic + Stochastic

- Stochastic just means unpredictable.
- the difference between the expected value and the observed value.
- series of errors, should look random

- deterministic component is the portion of the variation in the dependent variable that the independent variables explain.
- all of the explanatory power should reside here.

R-SQUARED

- Is a goodness-of-fit measure for linear regression models.
- Indicates the % of the variance in the dependent variable that the independent variables explain collectively.
- R-squared measures the strength of the relationship between your model and the dependent variable on a convenient 0 – 100% scale.
    - the R^2 is always going to be between $-\infty$ and 1
- *Small R-squared values are not always a problem, and high R-squared values are not necessarily good!*

R2 - INTERPRETATION

$$SST = \sum(y - \bar{y})^2$$

$$SSM = \sum(\hat{y} - \bar{y})^2$$

$$SSE = \sum(y - \hat{y})^2$$

$$R^2 = \frac{SSM}{SST}$$

R^2 will increase as we increase the number of features or complexity

Refer to python code
ML-LINREG-01-r2-adj-r2

EXERCISE

X	Y
1	8
2	8
3	8
4	8
5	8
6	8
7	8
8	8
9	8
10	8

R2 of this linear regression line is?

ADJUSTED R2

$$R^2_{adjusted} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

where

R^2 = sample R-square

p = Number of predictors

N = Total sample size.

INTERPRET P-VALUES AND COEFFICIENTS

- P-values and coefficients in regression analysis work together to tell which relationships in the model are statistically significant and the nature of those relationships.
- The coefficients describe the mathematical relationship between each independent variable and the dependent variable.
- The p-values for the coefficients indicate whether these relationships are statistically significant.
- *After fitting a regression model, check the residual plots first to be sure that we have unbiased estimates. After that, it's time to interpret the statistical output.*

BASICALLY …

- Regression analysis is a form of inferential statistics.
- The p-values help determine whether the relationships observed in the sample also exist in the larger population.
- The p-value for each independent variable tests the null hypothesis that the variable has no correlation with the dependent variable. (no effect)

$$p - value = \left[\begin{array}{c} \textit{if less than 0.05, then reject the Null hypothesis, that variables have no effect. Keep those variables in the model} \\ \hline \textit{if more than 0.05 then ACCEPT the Null hypothesis, that the variables have no effect. Remove these features} \end{array} \right]$$

SUMMARY

Acronym	Residual Operation?	Robust To Outliers?	Key Observations (pros)	Key Observations (cons)
MAE Mean Absolute Error	Absolute Value	Yes	• Mean magnitude of the residuals • larger errors will contribute linearly to the overall error. • small MAE suggests the model is great at prediction	• **does not** indicate if the model under or overshoots actual data). • large MAE suggests that your model may have trouble in certain areas

SUMMARY

Acronym	Residual Operation?	Robust To Outliers?	Key Observations (pros)	Key Observations (cons)

MSE
 	
Mean Squared Error
	Square	No	• Analogous to variance	• error grows **quadratically**
				
				• **Overestimating badness**
				  • Metric worsens with the square term in case of few outliers
				
				  • even a “perfect” model may have a high MSE in case of noisy data
				
				• **underestimate badness**
				  • **ALL the errors are small**

SUMMARY

Acronym	Residual Operation?	Robust To Outliers?	Key Observations (pros)	Key Observations (cons)
RMSE Root Mean Squared Error	Square	No	• a close relative of MSE, can compare models using MSE • Analogous to SD	• affected by the outliers

SUMMARY

Acronym	Residual Operation?	Robust To Outliers?	Key Observations (pros)	Key Observations (cons)
MAPE Mean Absolute Percentage Error	Absolute Value	Yes	• **percentage** equivalent of MAE	• undefined for data points where the value is 0 • can grow unexpectedly large if the actual values are exceptionally small themselves • biased towards predictions that are systematically less than the actual values

SUMMARY

Acronym	Full Name	Residual Operation?	Robust To Outliers?	Key Observations (pros)	Key Observations (cons)
MPE	Mean Percentage Error	N/A	Yes	• more negative or positive errors, this bias will show up in the MPE • Tells if our model systematically **underestimates** (more negative error) or **overestimates** (positive error).	Since positive and negative errors cancel out, cannot make any statements about how well the model predictions perform overall

PARAMETERS

```
sklearn.linear_model.LinearRegression(fit_intercept=True,
                                      normalize=False,
                                      copy_X=True,
                                      n_jobs=None)
```

Parameters	
fit_intercept	boolean, optional, default True whether to calculate the intercept for this model. If set to False, no intercept will be used in calculations
normalize	boolean, optional, default False If True, the regressors X will be normalized before regression by subtracting the mean and dividing by the l2-norm. If you wish to standardize, please use sklearn.preprocessing.StandardScaler before calling fit on an estimator with normalize=False.

IMPORTANT POINTS:

- There must be linear relationship between independent and dependent variables
- Multiple regression suffers from multicollinearity, autocorrelation, heteroskedasticity.
- Linear Regression is very sensitive to Outliers. It can terribly affect the regression line and eventually the forecasted values.
- Multicollinearity can increase the variance of the coefficient estimates and make the estimates very sensitive to minor changes in the model. The result is that the coefficient estimates are unstable
- In case of multiple independent variables, go with *forward selection, backward elimination and step wise approach for selection of most significant independent variables.*

WHEN TO USE LIN REG

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

where,

$\beta_1, \beta_2, \ldots, \beta_p$ are the regression slope coefficients,

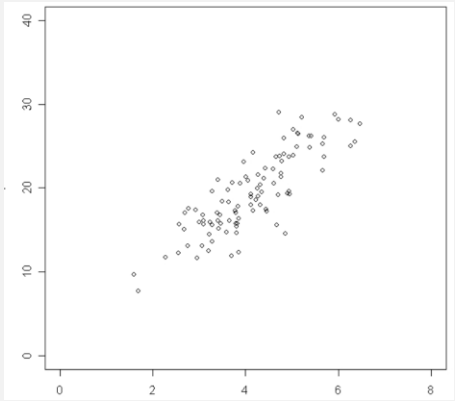
$X_{i1}, X_{i2}, \ldots, X_{ip}$ are the corresponding values of the X-variables,

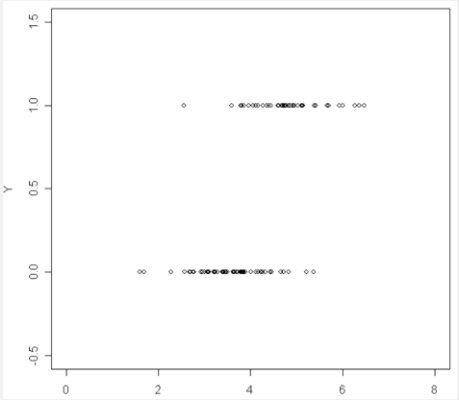
α is the intercept, and

ε_i is the random error term.

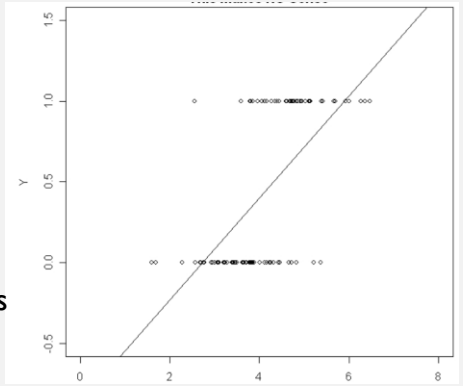
the error term Epsilon has a normal distribution with mean 0 and standard deviation

This means the output y, must be continuous variable (i.e., one that takes values on an interval), not a binary variable





- The least-squares regression line gives predictions that make no sense.
- For example, for an X value of 8, the least-square regression line predicts that Y will be above 1.5.
- But Y can only take on values of 0 and 1.



ASSUMPTIONS

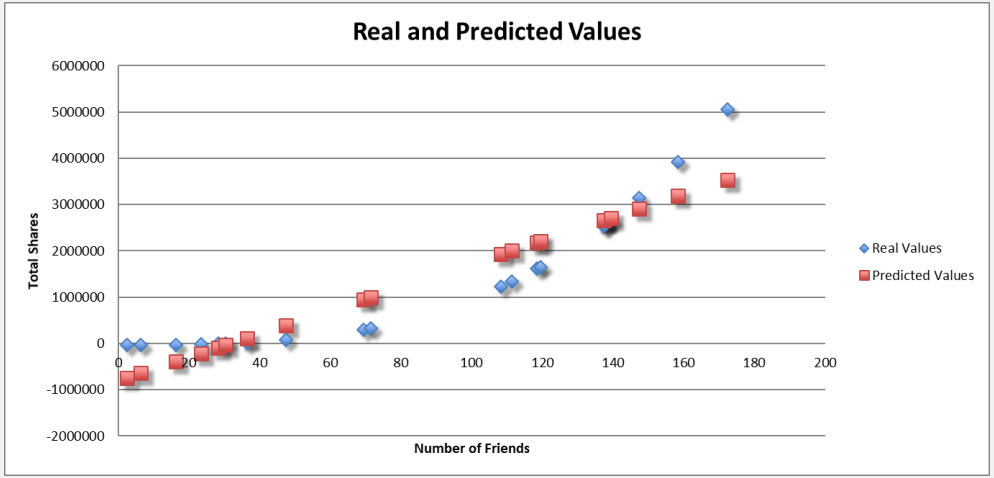
Criteria	
Number of observations	the observation-to-Independent Variables (IVs) ratio should ideally be 20:1; that is 20 cases for every IV in the model. For 10 IV, total sample size min 200
Accuracy of data	Pandas (describe method) at least check the minimum and maximum value for each variable to ensure that all values for each variable are valid.
Missing data	Consider imputation (In case of few values being missing)
Outliers	• An outlier is often operationally defined as a value that is at least 3 SD above or below the mean. • delete those cases. • recode the value
Linearity	• Use bivariate scatterplots to check. • Look for residual vs fitted value plots
Heteroscedasticity of residuals	• One of the important assumptions of linear regression is that, there should be no heteroscedasticity of residuals. In simpler terms, this means that the variance of residuals should not increase with fitted values of response variable.

ASSUMPTIONS

Criteria	
Normality	construct histograms and "look" at the data to see its distribution.
Multicollinearity	• Multicollinearity is a condition in which the IVs are very highly correlated (.90 or greater) • Ordinary Least square method cannot be that effective. • deleting one of the two variables (after careful analysis) • Statistically, regression coefficients is calculated through matrix inversion. if multicollinearity exists the inversion is unstable.
Transformations	• one or more variables are not normally distributed, you might want to transform them • harder to interpret the analysis

WHEN, WHY AND HOW SHOULD USE LINEAR REGRESSION

- The relationship between the variables is linear.
  - variance in the residuals (the difference in the real and predicted values) is more or less constant.
  - The residuals are independent, meaning the residuals are distributed randomly and not influenced by the residuals in previous observations.
  - The residuals are normally distributed. This assumption means the



USE CASES

- Sales of a product; pricing, performance, and risk parameters
- Generating insights on consumer behavior, profitability, and other business factors
- Evaluation of trends; making estimates, and forecasts
- Determining marketing effectiveness, pricing, and promotions on sales of a product
- Assessment of risk in financial services and insurance domain
- Studying engine performance from test data in automobiles
- Calculating causal relationships between parameters in biological systems
- Conducting market research studies and customer survey results analysis
- Astronomical data analysis
- Predicting house prices with the increase in sizes of houses
- stock trading, video games, sports betting, and flight time prediction.

GEOMETRIC INTUITION

- Models with 2 predictor variables (say x_1 and x_2) and a response variable y can be understood as a two-dimensional surface in space.

- The observations are points in space and the surface is "fitted" to best approximate the observations.

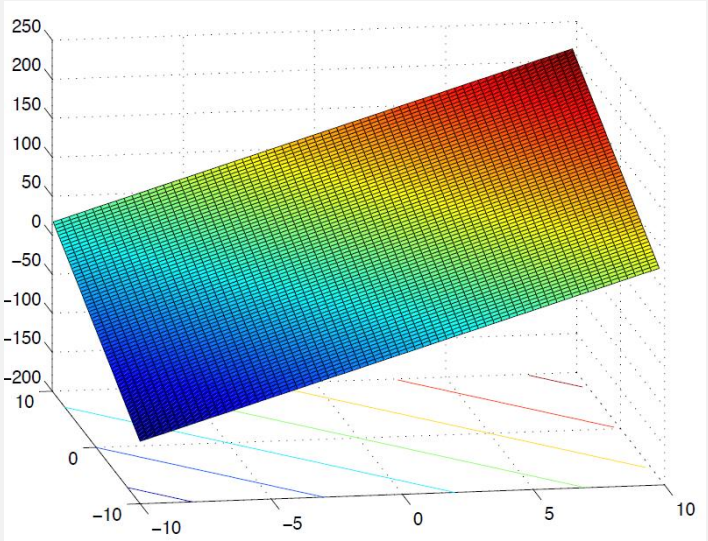
- Simplest multiple regression model for 2 predictor variables is

  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

- The surface that corresponds to the model

  $y = 50 + 10x_1 + 7x_2$ looks like this.

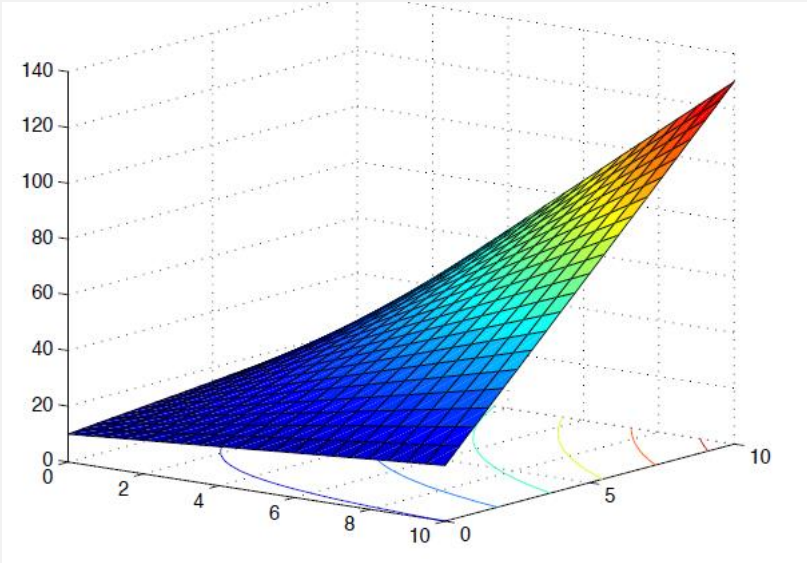
- It is a plane in R3 with different slopes in x_1 and x_2



GEOMETRIC INTUITION

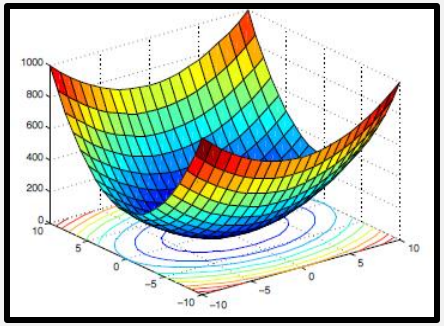
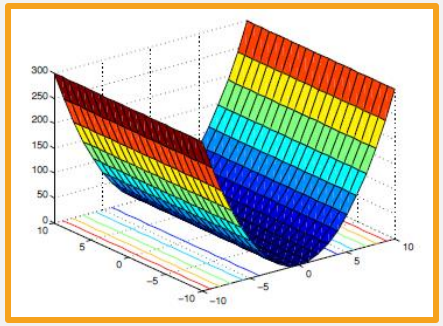
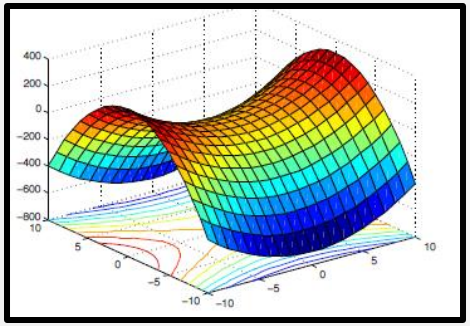
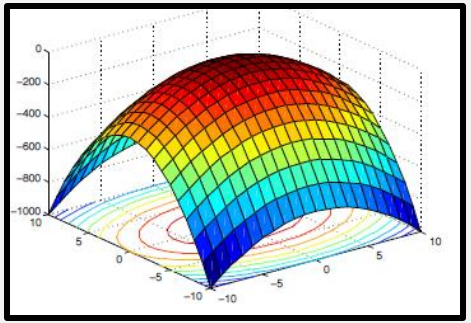
- For a simple linear model with two predictor variables and an interaction term, the surface is no longer flat but curved.

- $y = 10 + x_1 + x_2 + x_1x_2$



GEOMETRIC INTUITION

- Polynomial regression models with 2 predictor variables and interaction terms are quadratic forms.
- Their surfaces can have many different shapes depending on the values of the model parameters with the contour lines being either parallel lines, parabolas or ellipses.
- $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_2^1 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$

QS

Linear Regression is a supervised machine learning algorithm.

A) TRUE
B) FALSE

Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B

QS

Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?

A) AUC-ROC
B) Accuracy
C) Logloss
D) Mean-Squared-Error

Which of the following is true about Residuals ?

A) Lower is better
B) Higher is better
C) A or B depend on the situation
D) None of these

QS

Suppose that we have N independent variables (X1,X2… Xn) and dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data.

You found that correlation coefficient for one of it's variable(Say X1) with Y is -0.95.

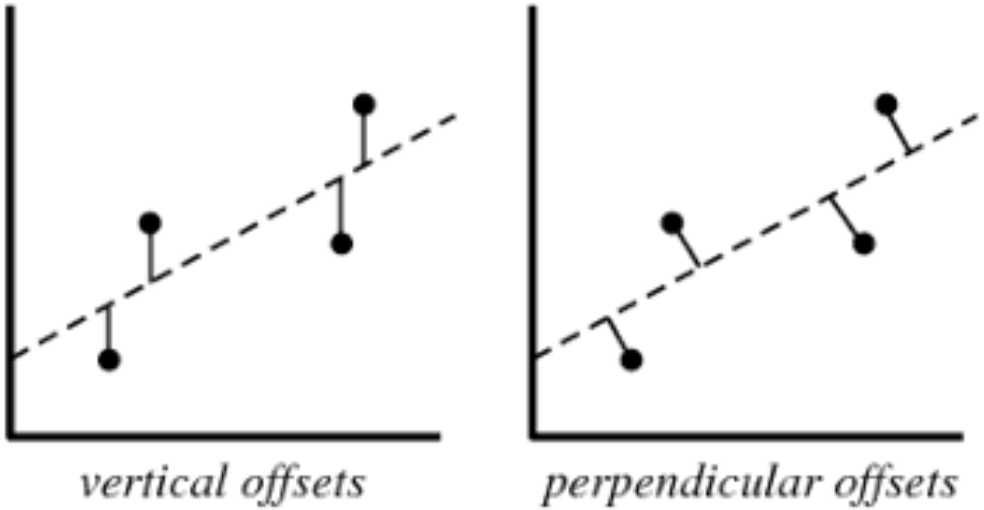
Which of the following is true for X1?

A) Relation between the X1 and Y is weak
B) Relation between the X1 and Y is strong
C) Relation between the X1 and Y is neutral
D) Correlation can't judge the relationship

QS

Which of the following offsets, do we use in linear regression's least square line fit? Suppose horizontal axis is independent variable and vertical axis is dependent variable.

A) Vertical offset
B) Perpendicular offset
C) Both, depending on the situation
D) None of above



QS

Which of the following statement is true about outliers in Linear regression?

A) Linear regression is sensitive to outliers
B) Linear regression is not sensitive to outliers
C) Can't say
D) None of these

Suppose you plotted a scatter plot between the residuals and predicted values in linear regression and you found that there is a relationship between them. Which of the following conclusion do you make about this situation?

A) Since the there is a relationship means our model is not good
B) Since the there is a relationship means our model is good
C) Can't say
D) None of these

QS

What will happen when you fit degree 4 polynomial in linear regression?

A) There are high chances that degree 4 polynomial will over fit the data
B) There are high chances that degree 4 polynomial will under fit the data
C) Can't say
D) None of these

Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error?

A) Increase
B) Decrease
C) Remain constant
D) Can't Say

QS 2

At a .01 level of significance is there sufficient evidence to conclude that the number of books sold is related to the number of registered students in a straight-line manner?

H_0: The number of students registered and the number of books sold are not correlated
H_a: The number of students registered and the number of books sold are correlated

Decision Rule: Accept Ha if the calculated p-value < .01

Test Statistic: r = the Pearson coefficient of correlation
Calculations : r = 0.8997,
r^2 = 0.80
p-value < 0.0001 < .01 (α)

Interpretation: At the .01 level of significance , the probability that observed values can be found (H_O = true) is very very less than given α . Reject H_O. Hence there is strong relation between number of students and books

QS 3

Carefully explain what the p-value found, means.

H_0: The number of students registered and the number of books sold are not correlated
H_a: The number of students registered and the number of books sold are correlated

Decision Rule: Accept Ha if the calculated p-value < .01

Test Statistic: r = the Pearson coefficient of correlation
Calculations : r = 0.8997,
r^2 = 0.80
p-value < 0.0001 < .01 (α)

Interpretation: At the .01 level of significance , the probability that observed values can be found (H_O = true) is very very less than given α . Reject H_O. Hence there is strong relation between number of students and books

QS 4

interpret the strength of the straight-line relationship.

r^2 = .809 (80.9%).

Coefficient of determination

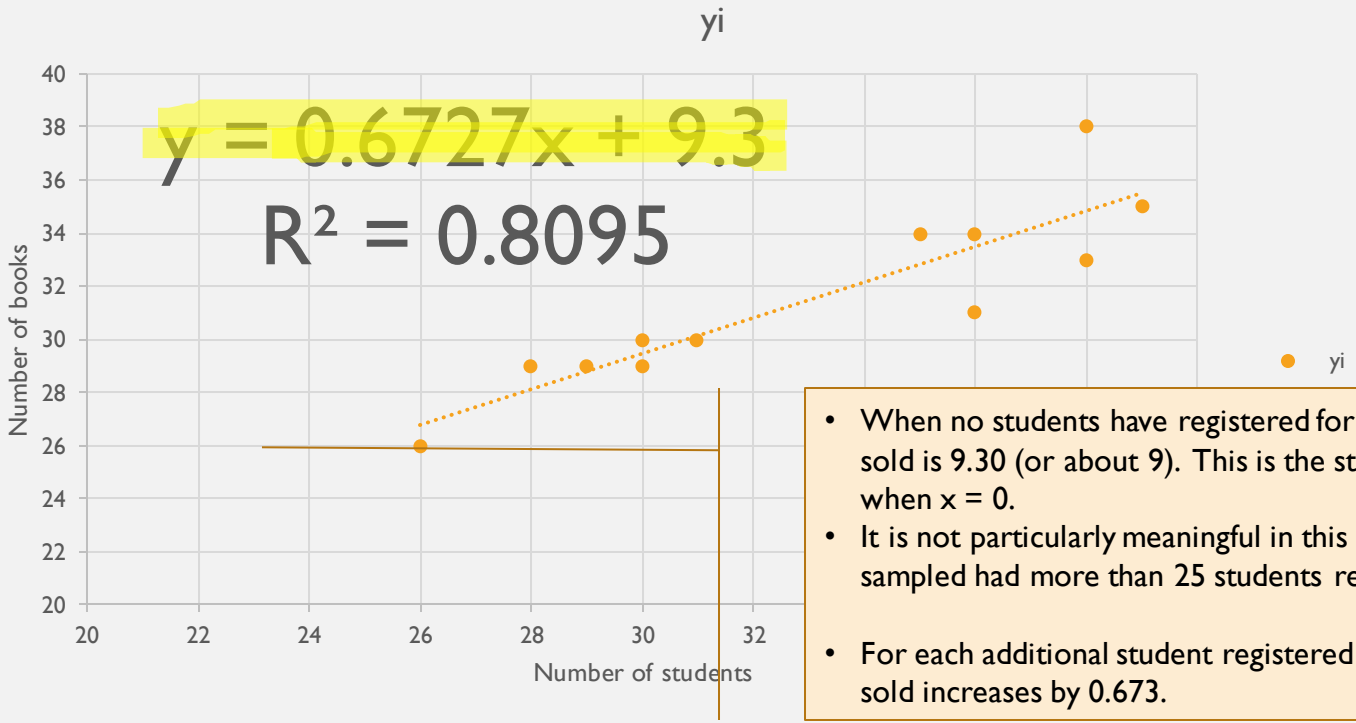
80.9% of the variability in the number of books sold is explained by the straight-line relationship with the number of registered students.

19.1% of this variability is unexplained, and due to error.

This relationship is quite strong.

QS 5

Give the regression equation, and interpret the coefficients in terms of this problem..



- When no students have registered for a course, the number of books sold is 9.30 (or about 9). This is the starting point of the straight-line when x = 0.
- It is not particularly meaningful in this problem since all the classes sampled had more than 25 students registered.

- For each additional student registered for a course, the number of books sold increases by 0.673.

QS 6

predict the number of books that would be sold in a semester when 30 students have registered. Use 95% confidence...

Since 30 students is within the range of the sampled number of students, it is appropriate to make this prediction.

From excel, the calculated prediction interval is (25.865078,33.09856).

95% confident that for a course that has 30 students registered the bookstore will sell between 25.9 and 33.1 books.

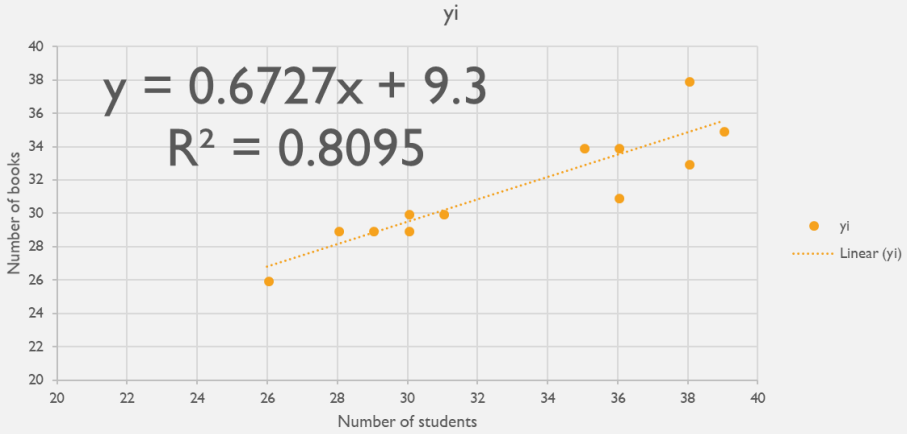
Y-pred = 29.48181818

QS 7

If appropriate, predict the number of books that would be sold in a semester when 5 students have registered. Use 95% confidence.

Since 5 students is not within the range of the sampled number of students, it is not appropriate to use the regression equation to make this prediction.

We do not know if the straight-line model would fit data at this point, and we should not extrapolate.



P-VALUES

- Hypothesis tests are used to test the validity of a claim that is made about a population.
- This claim that's on trial, in essence, is called the null hypothesis. (H_o)

> The null hypothesis is usually an hypothesis of "no difference" e.g. no difference between blood pressures in group A and group B.

- The alternative hypothesis is the one you would believe if the null hypothesis is concluded to be untrue. (H_a)
- The p-value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H_o) of a study question is true
- All hypothesis tests ultimately use a p-value to weigh the strength of the evidence (what the data are telling about the population).

P-VALUES

- The p-value is a number between 0 and 1 and interpreted in the following way:

1. A small p-value (typically ≤ 0.05 or $\alpha = 5\% = .05$) indicates strong evidence against the null hypothesis, so you reject the null hypothesis.

2. A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis.

3. p-values very close to the cutoff (0.05) are considered to be marginal (could go either way).

EXAMPLE

For example, suppose a pizza place claims their delivery times are 30 minutes or less on average but you think it's more than that.

You conduct a hypothesis test with some sample data.

- null hypothesis, H_o = mean delivery time is 30 minutes max
- alternative hypothesis (H_a) = mean time is greater than 30 minutes.

After the test, the p-value turns out to be 0.001, which is much less than 0.05. (α)

It means there is only a probability of 0.001 that we will find observed instances when H_o = true. Hence reject H_o as it is << α