



# Regression

Del 1



# Agenda

- Regression
  - Vad
  - Enkel linjär
    - Delar
    - Variationer - exempel
    - Hur
    - Antaganden
    - Kod
    - Övning
  - Flervariabel linjär
  - Polynom( teori)

# Regression

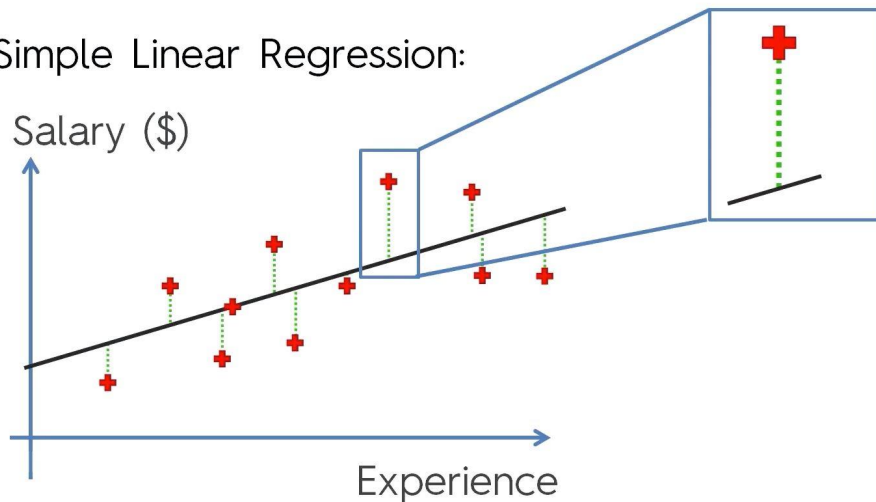
*"Regression analysis is a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other."*

- [alchemer.com...](#)

# Regression

- Finns det ett förhållande mellan de här två variabler?
- Hur starkt är sambandet?
- Kan använda flera variabler.
- Förutspå nya värden.

Simple Linear Regression:



# Enkel linjär

- Enkel - endast två variabler
- Tittar på **förhållandet** mellan dem.
- Vad blir y vid olika x värden  $y = f(x)$
- y är beroende av x (dependent variable)
- x är oberoende av y (independent)
- Linjär - värdet av y "följer" värdet av x



$$y = b + m * x$$

$$y = \beta_0 + \beta_1 * x$$

Ref: Länk 1.

**RESTORY...**

# Enkel Linjär

Simple  
Linear  
Regression

$$y = b + m * x$$

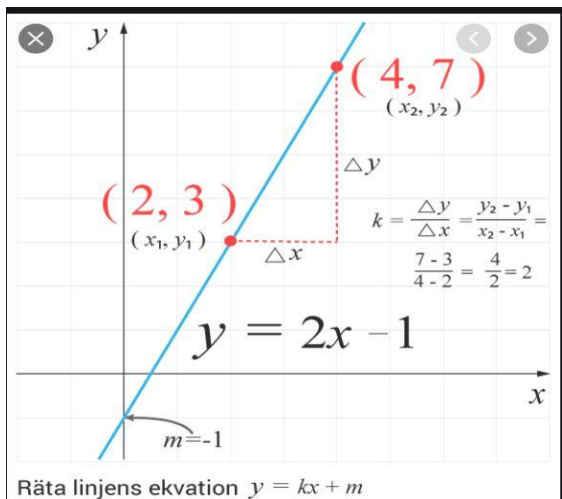
Diagram illustrating the Simple Linear Regression equation  $y = b + m * x$  with labels and arrows:

- Dependent variable (DV)**: Points to  $y$ .
- Coefficient**: Points to  $m$ .
- Independent variable (IV)**: Points to  $x$ .

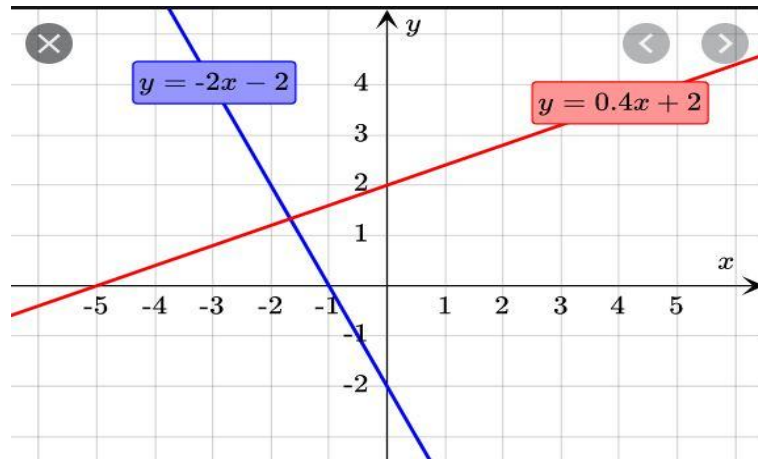
- Vad representerar konstanterna  $m$  och  $b$ ?
- **$m$**  kallas **riktningskoefficient** och bestämmer linjens lutning, aka “unit change”
- när  **$m$**  är **positivt** lutar linjen **uppåt**
- när  **$m$**  är **negativt** lutar linjen **neråt**
- när storleken på  **$m$**  **växer** (negativt eller positivt) blir linjen **brantare**
- **$b$**  representerar linjens **skärningspunkt med  $y$  axeln** (intercept)
- alla linjer med **samma  $m$**  värde är **parallella**

# Exempel - variationer

$\Delta = \delta = \text{delta}$  skillnaden

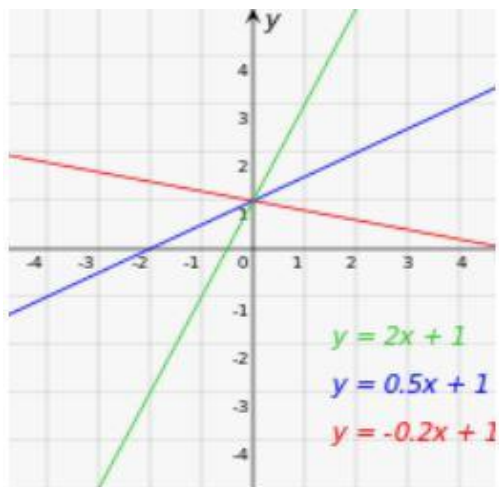


Positivt vs. negativt m och b

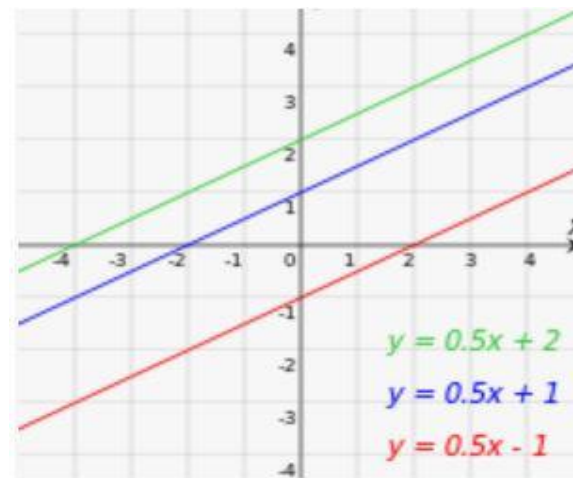


# Exempel - m och b variationer

Samma b, olika m



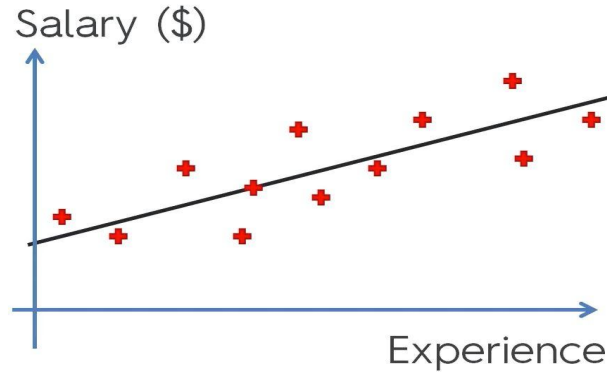
Samma m, olika b.





# Exempel

Simple Linear Regression:



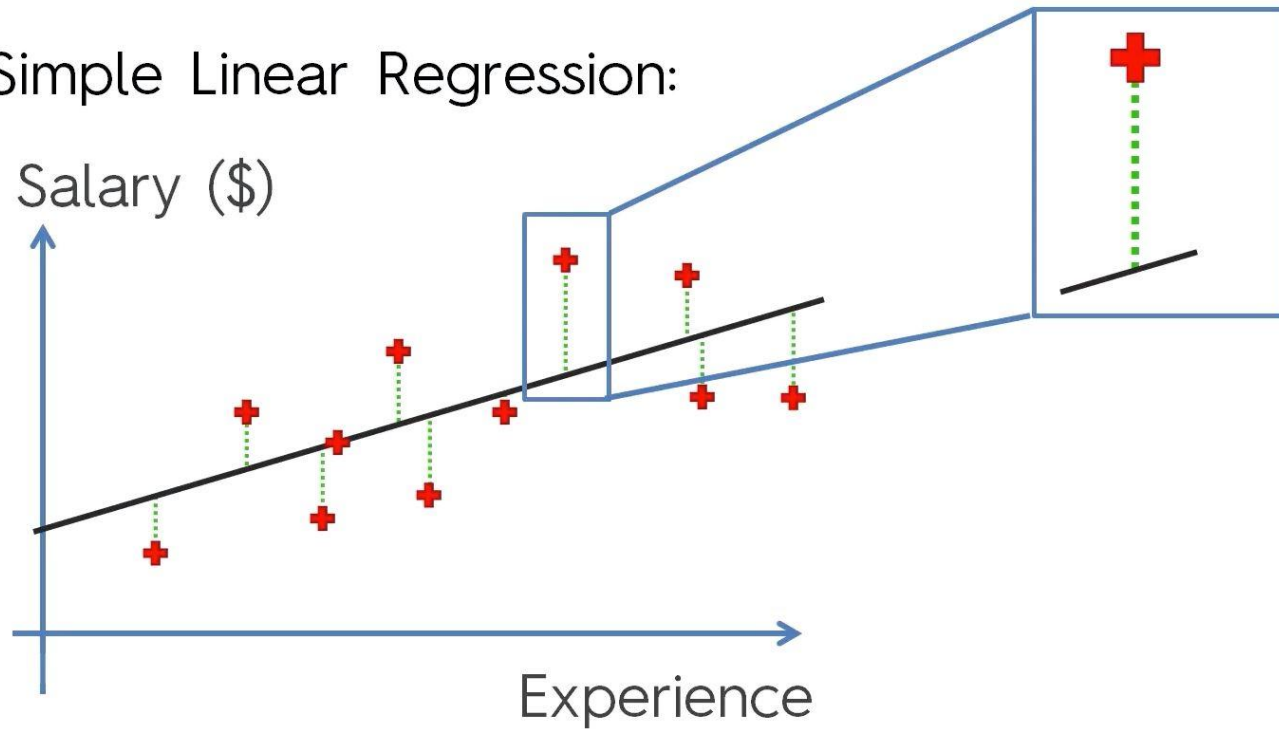
$$y = b + m * x$$



$$\text{Salary} = b + m * \text{Experience}$$

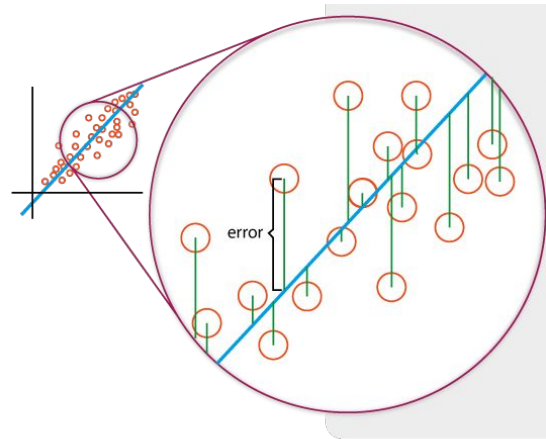
# Regression

Simple Linear Regression:



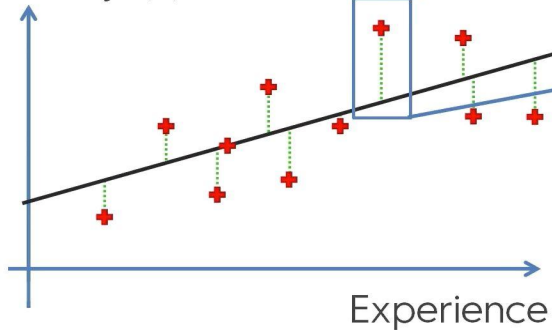
# Regression

- Skillnaden (residual = error)
- Summera alla
- Upphöjt (squared) - hantera "+/-"
- Vill minimera
- Stora skillnader har större påverkan
- Små värden - finjustering



Simple Linear Regression:

Salary (\$)



Observerat

Modellerat

$$\text{SUM } (y - \hat{y})^2 \rightarrow \min$$

**RESTORY...**

# Linjär Regression - Antaganden

1. Linearity
  - a. Det finns en linjär relation mellan den oberoende och beroende variabeln
2. Homoscedasticity
  - a. Homogenitet av variansen
  - b. Standardavvikelsen för alla normalfördelade grupper av Y-värden är lika.
3. Multivariate normality
4. Independence of errors
  - a. Felen (residuals) är oberoende (av  $y$ ).
5. **Lack** of multicollinearity
  - a. "One predictor variable in a multiple regression model cannot be linearly predicted from the others with a substantial degree of accuracy. "

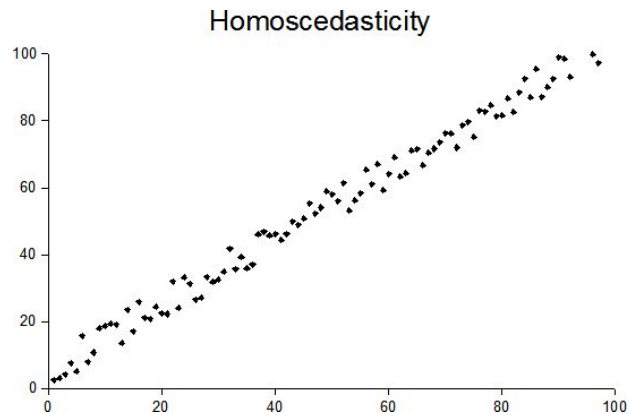


Fig: By Q9 at the English-language Wikipedia, CC BY-SA 3.0,  
<https://commons.wikimedia.org/w/index.php?curid=16464531>



Kod!



**RESTORY...**

# Kod

- pandas
- matplotlib
  - pyplot
- sklearn
  - model\_selection
    - train test split
      - **random\_state int, default=None** - Controls the shuffling applied to the data before applying the split. Pass an int for reproducible output across multiple function calls.
- linear\_model
  - LinearRegression
    - fit
      - x\_vals
      - y\_vals
    - predict
      - new x\_vals
    - coef\_
    - intercept\_

# Övning

- [Ladda ned diamonds data från Kaggle](#)
- Gör en linjär regression av
  - pris och "cut"
  - pris och "clarity"
  - pris och färg
- Gör lite snygga grafer för att visa upp ditt resultat
- Testa med annan data



# Linjär Regression med flera variabler

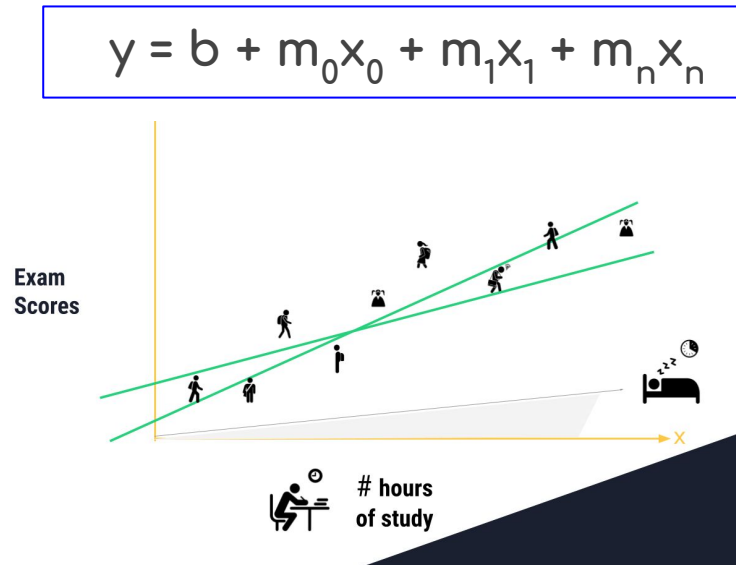


# Flervariabel Linjär Regression

y är nu beroende av två, eller fler, variabler  
(years of experience AND years of education)

Varje ny variabel är också oberoende.

Har sin egen riktningskoefficient.



<https://images.app.goo.gl/ZuAVS1F2JuiTXN1B8>

RESTORY...

# Flervariabel LR

Simple  
Linear  
Regression

$$y = b_0 + b_1 * x_1$$

Multiple  
Linear  
Regression

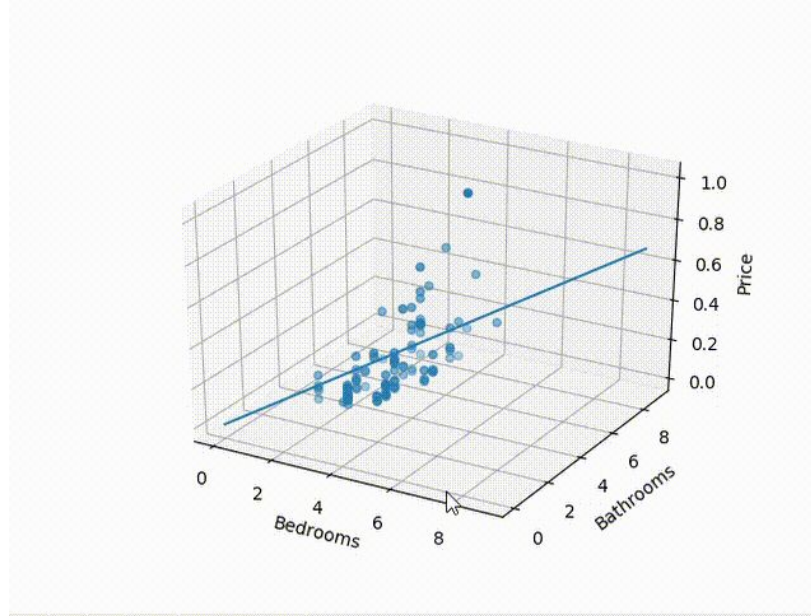
Dependent variable (DV)      Independent variables (IVs)

The diagram shows the equation  $y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$ . Green arrows point from labels to parts of the equation: 'Dependent variable (DV)' points to 'y'; 'Independent variables (IVs)' points to 'x<sub>1</sub>' and 'x<sub>2</sub>'; 'Constant' points to 'b<sub>0</sub>'; and 'Coefficients' points to 'b<sub>1</sub>', 'b<sub>2</sub>', and 'b<sub>n</sub>'.

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Constant      Coefficients

# Flervariabler - två oberoende variabler - 3D



bedrooms  
bathrooms

price

# The dummy variable trap

Ordinal to numerical

Third can be inferred by the absence of the other two.

sklearn model is advanced and will figure that out

# Övning

- Ladda ned Iris data
  - Omvandla nominell data till numeriska kategorier
  - Gör en flervariabelregression för att förutspå Sepal längden.
- 
- [Ladda ned Pima indians diabetes data](#)
  - Gör en flervariabelregression.



# Polynom



**RESTOR\...**

# Polynom Linjär Regression

Simple  
Linear  
Regression

$$y = b_0 + b_1x_1$$

Multiple  
Linear  
Regression

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Polynomial  
Linear  
Regression

$$y = b_0 + b_1x_1 + b_2x_1^2 + \dots + b_nx_1^n$$

# Polynom LR

$$y = b + m_0x + m_1x^2 + \dots + m_nx^n$$

y är beroende av x

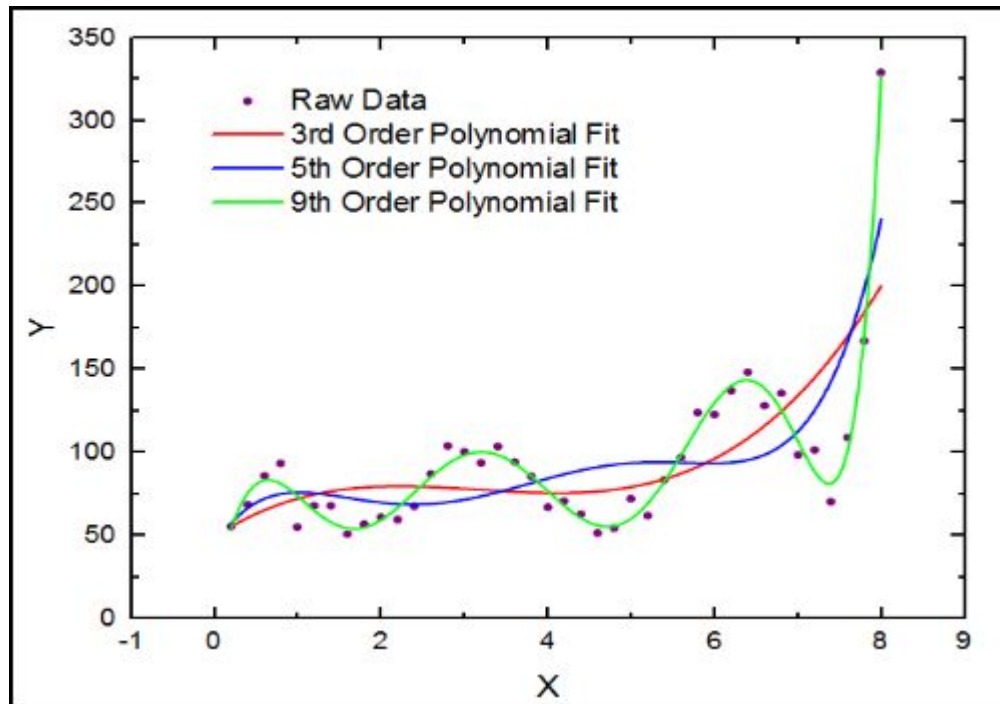
x är oberoende

förhållandet mellan dem som en  
n-graders polynom av x  
(relationship between them as a n-th  
degree polynomial of x)

högre grad -> mer komplex modell

Kan modellera mer komplexa  
förhållanden.

Antalet parametrar går upp.

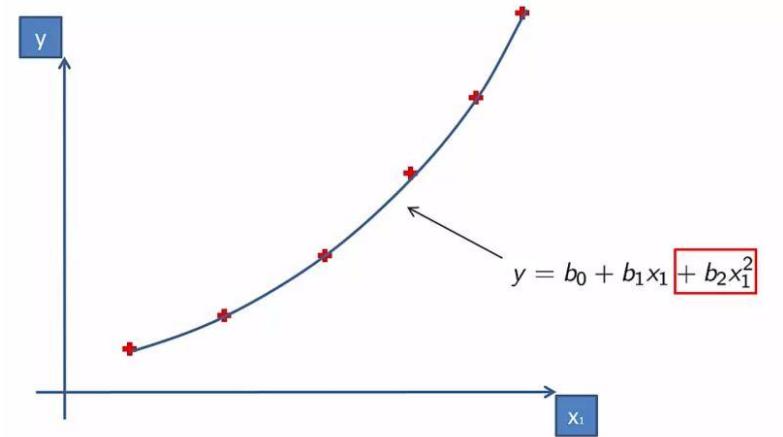
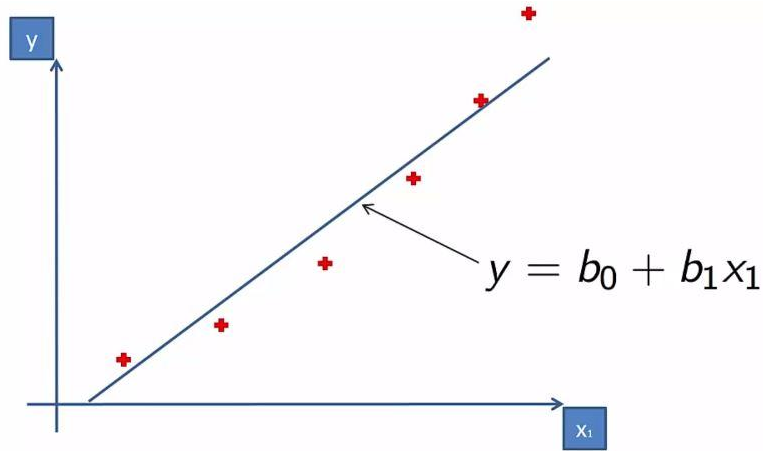


<https://images.app.goo.gl/qriDKw1nLNMr2fmf9>

**RESTORY...**



# Polynom regression - när behövs det



RESTOR\...

# Länkar

- [Linear regression basics for absolute beginners](#)
- [Linear regression clearly explained - video](#)
- [Examples of Linear regression and sklearn](#)
- [Regression och korrelation](#)
- [Heteroscedasticity](#)
- [Regression - olika typer](#)
- [An inspiring journey](#)