



Artificial Learning Models

Lecture 2 : Linear Regression

By : Dr. Lamri SAYAD

2024

Agenda

Problems for Today

- How many views/subscribers can I get on my Youtube channel ?



artificial intelligence



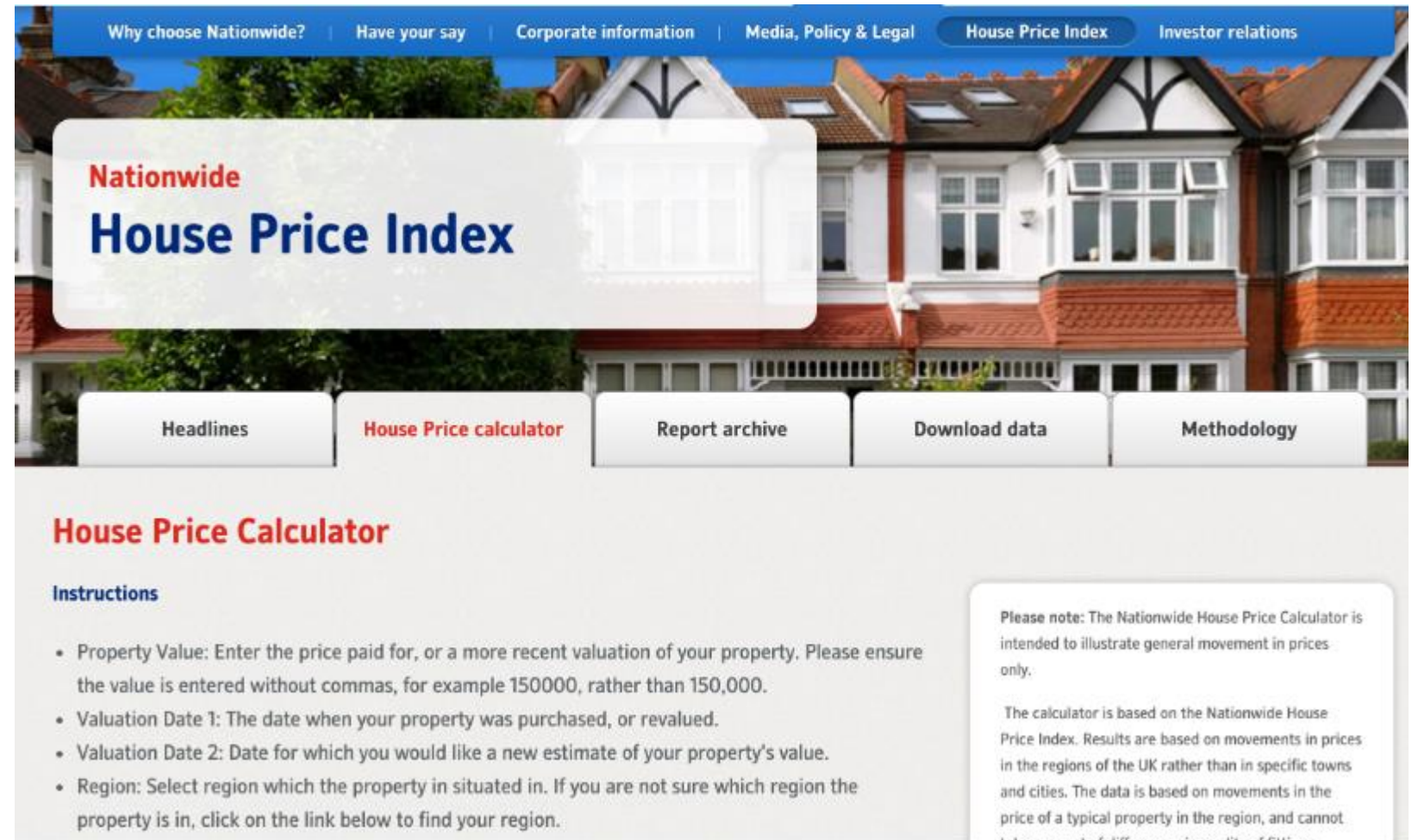
الآلة ضد الإنسان: هل سيصبح الإنسان ضحية الذكاء الاصطناعي؟

76 k vues • il y a 3 mois



Problems for Today

- Goal: Predict the price of the house



The screenshot shows the Nationwide House Price Index website. The header includes navigation links: "Why choose Nationwide?", "Have your say", "Corporate information", "Media, Policy & Legal", "House Price Index" (highlighted), and "Investor relations". A large white box in the center contains the text "Nationwide House Price Index". Below this, a row of buttons includes "Headlines", "House Price calculator" (highlighted), "Report archive", "Download data", and "Methodology". The main content area is titled "House Price Calculator" and includes a section for "Instructions" with a bulleted list of input requirements. A "Please note" box on the right explains the calculator's purpose and limitations.

Nationwide House Price Index

Headlines | **House Price calculator** | Report archive | Download data | Methodology

House Price Calculator

Instructions

- Property Value: Enter the price paid for, or a more recent valuation of your property. Please ensure the value is entered without commas, for example 150000, rather than 150,000.
- Valuation Date 1: The date when your property was purchased, or revalued.
- Valuation Date 2: Date for which you would like a new estimate of your property's value.
- Region: Select region which the property is situated in. If you are not sure which region the property is in, click on the link below to find your region.

Please note: The Nationwide House Price Calculator is intended to illustrate general movement in prices only.

The calculator is based on the Nationwide House Price Index. Results are based on movements in prices in the regions of the UK rather than in specific towns and cities. The data is based on movements in the price of a typical property in the region, and cannot take account of differences in quality of fittings.

What do all these problems have in common?

- Continuous outputs
- Predicting continuous outputs is called regression

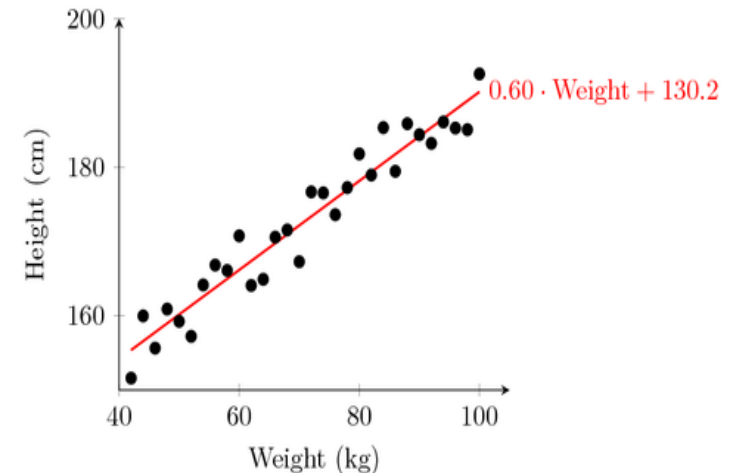


Linear regression :

- Continuous outputs
- Simple model (linear)

What is Linear Regression ?

- Used to predict the relationship between two variables.
- Fit a line to a data set of observations
 - It assumes a linear relationship between the independent variable and the dependent variable
- Use this line to predict unobserved values
- Key Questions: How to determine this line ?
 - How do we parametrize the model?
 - What loss (objective) function should we use to judge the fit?
 - How do we optimize fit to unseen test data (generalization)?



Regression

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression

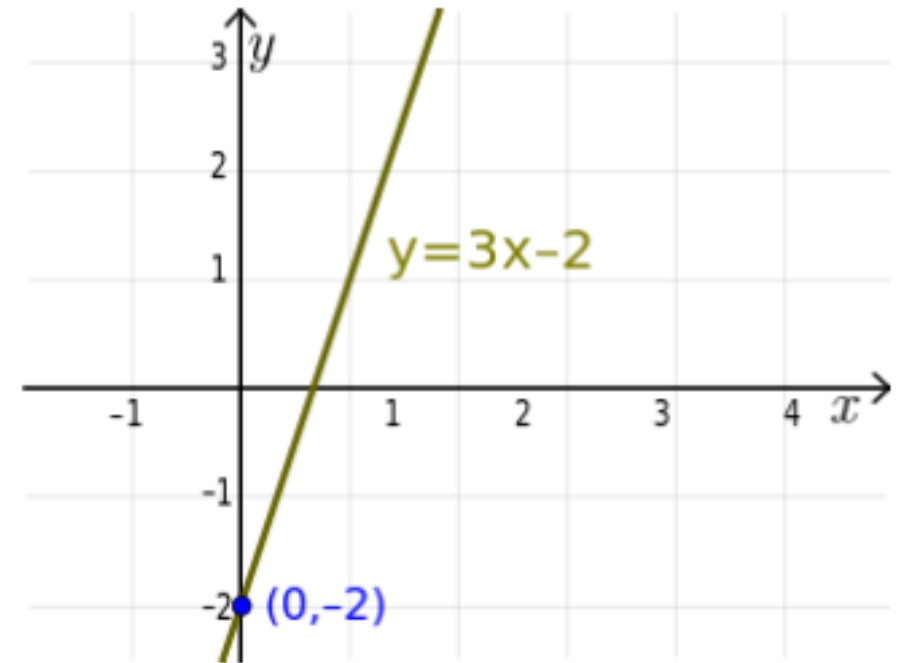
Simple Linear Regression

Simple Linear Regression

- One independent variable and one dependent variable.
- The model estimates the slope and intercept of the line of best fit
 - The relationship between the variables.
- If there is a single input variable X (independent variable), such linear regression is called ***simple linear regression***

How does it work?

- Using a Cost function (To be minimized)
- Linear model : $\hat{y} = \phi(x) = w_1 \cdot x + w_0$
- The slope w_1 is the correlation between the two variables times the standard deviation in Y, all divided by the standard deviation in X.
- The intercept w_0 is the mean of Y minus the slope times the mean of X
- But Python will do all that for you.



Cost Function for Linear Regression

- Helps to find the optimal values of the slope and the intercept
- Provides the fit line for the data points
- Commonly used cost functions:
 - **Mean Squared Error (MSE):** average of squared error that occurred between the $y_{\text{predicted}}$ and y_i .

$$MSE = \ell(w) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - (w_1 \cdot x + w_0))^2$$

- **Least Squared Error LSE**

$$LSE = \ell(w) = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - (w_1 \cdot x + w_0))^2$$

- **Goal:** find the best values of w_0 and w_1 that minimizes the cost function

How do we obtain $w = (w_0, w_1)$?

Obtain w_0 and w_1

- Optimizing across training set
 - Gradient descent
- Analytical solution (in some cases)
 - in the case of linear least-squares regression
- How?

$$w = (X^T X)^{-1} X^T Y$$

Evaluation Metrics for Linear Regression

- Evaluation metrics used to assess the strength of any linear regression model
- Measure of how well the observed outputs are being generated by the model.
- The most used metrics are:
 - Coefficient of Determination or R-Squared (R^2)
 - Root Mean Squared Error (RSME) and Residual Standard Error (RSE)

R-squared

- It ranges between 0 and 1
- The higher the value of R-squared, the better the model fits the data.
 - 0 is bad (none of the variance is captured),
 - 1 is good (all of the variance is captured).

$$R^2 = 1 - (RSS/TSS)$$

$$RSS = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

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

RSS : Residual sum of Squares, **TSS** : Total Sum of Squares

$\bar{y}$ is the mean of the sample data points.

Root Mean Squared Error

- It specifies how close the observed data points are to the predicted values.
- Mathematically

$$RMSE = \sqrt{\frac{RSS}{n}} = \sqrt{\sum_{i=1}^n (y_i^{Actual} - y_i^{Predicted})^2 / n}$$

- R-squared is a better measure than RSME.

Multiple Regression

Multiple Regression

- What if more than one variable influences the one you're interested in?
- Example: predicting a price for a car based on its many attributes (body style, color, number of seats, ...)
- Technique to understand the relationship between a *single* dependent variable and *multiple* independent variables.

$$Y = w_0 + w_1X_1 + w_2X_2 + \dots + w_pX_p = W^T X$$

Still uses least squares

- We just end up with coefficients for each factor.
 - For example, $price = \alpha + \beta_1 \cdot \text{mileage} + \beta_2 \cdot \text{age} + \beta_3 \cdot \text{doors}$
 - These coefficients imply how important each factor is (if the data is all normalized!)
- Can still measure fit with r-squared
- Need to assume the different factors are not themselves dependent on each other.

What if our linear model is not good? How can we create a more complicated model?

- How do we know that a linear regression model is the best choice?
- What other types of regression are there?
 - There are many other types.
 - The linear model is by far the simplest, but it is not the only choice.

Nonlinear Regression

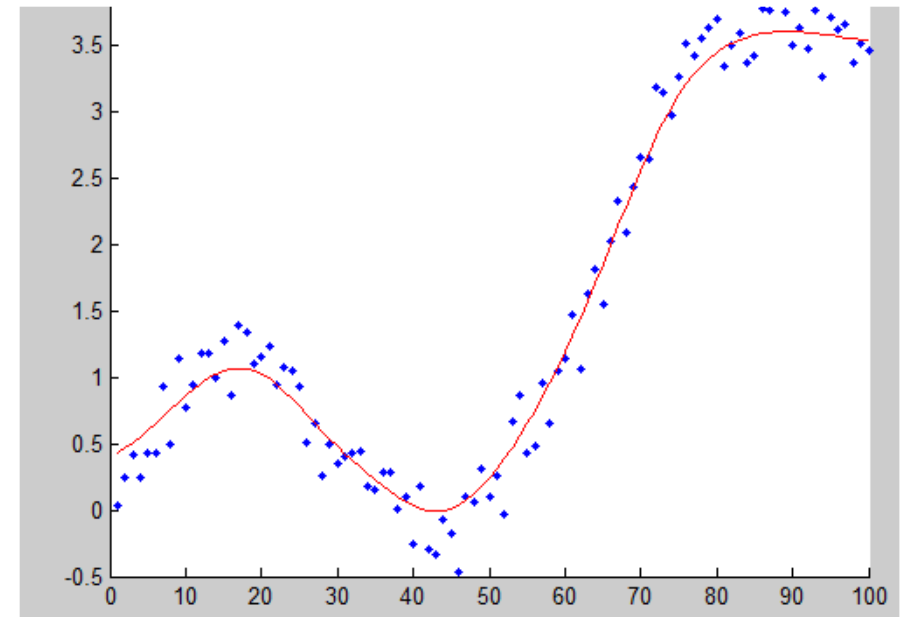
- Nonlinear Regression (Polynomial regression):
 - Quadratic regression: $\hat{y} = ax^2 + bx + c$.
 - Cubic regression: $\hat{y} = ax^3 + bx^2 + cx + d$.
 - Quartic regression: $\hat{y} = ax^4 + bx^3 + cx^2 + dx + e$
- Other Nonlinear Regression (Not polynomial)
 - Logarithmic regression: $\hat{y} = a + b \ln(x)$.
 - Exponential regression: $\hat{y} = ab^x$
 - Power regression: $\hat{y} = ax^b$
 - Logistic regression: $\hat{y} = \frac{c}{1+ae^{-bx}}$
 - ...

Polynomial Regression

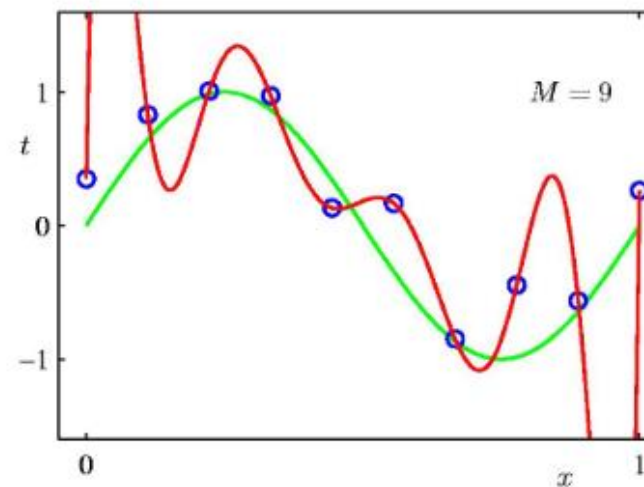
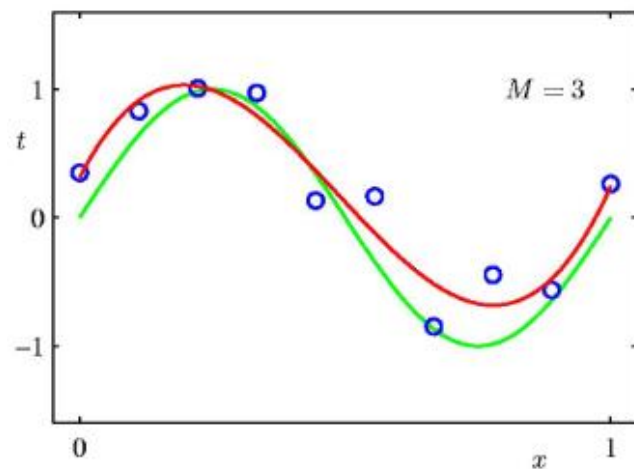
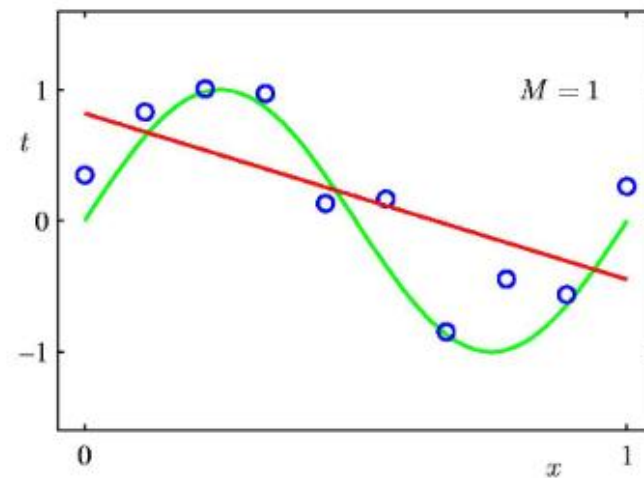
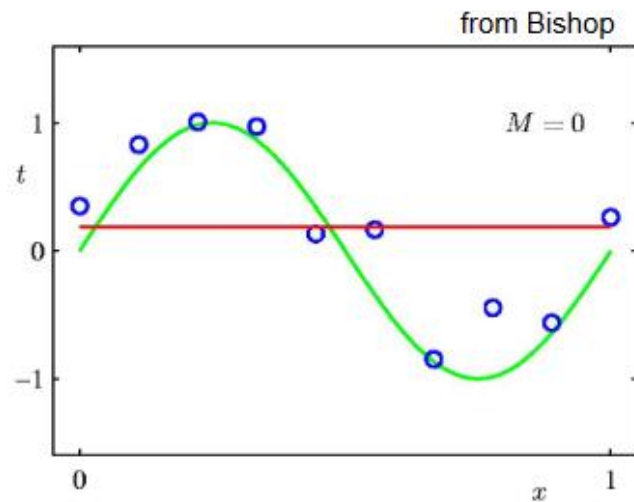
Why limit ourselves to straight lines?

- Not all relationships are linear.
- Linear formula: $y = mx + b$
 - This is a “first order” or “first degree” polynomial, as the power of x is 1
- Second order polynomial: $y = ax^2 + bx + c$
- Third order: $y = ax^3 + bx^2 + cx + d$
- Higher orders produce more complex curves.

$$y(x, \mathbf{w}) = w_0 + \sum_{j=1}^M w_j x^j$$



Which Fit is Best?

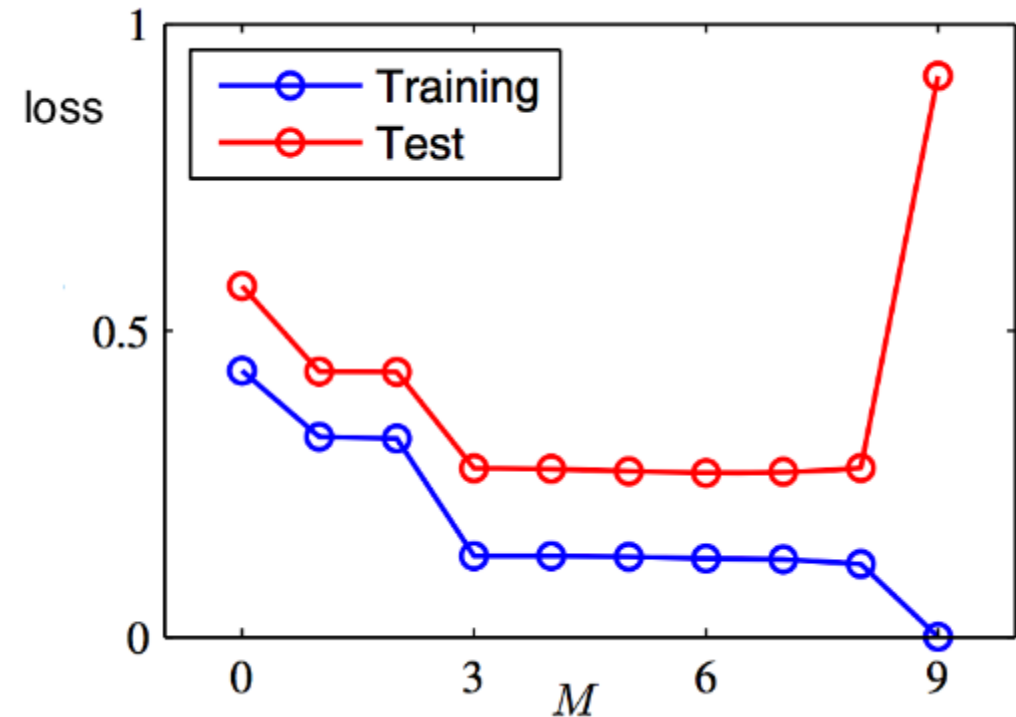


Beware overfitting

- Don't use more degrees than you need
- Visualize your data first to see how complex of a curve there might really be
- Visualize the fit is your curve going out of its way to accommodate outliers?
- A high r^2 simply means your curve fits your training data well; but it may not be a good predictor.
- Later we'll talk about more principled ways to detect overfitting (train/test)

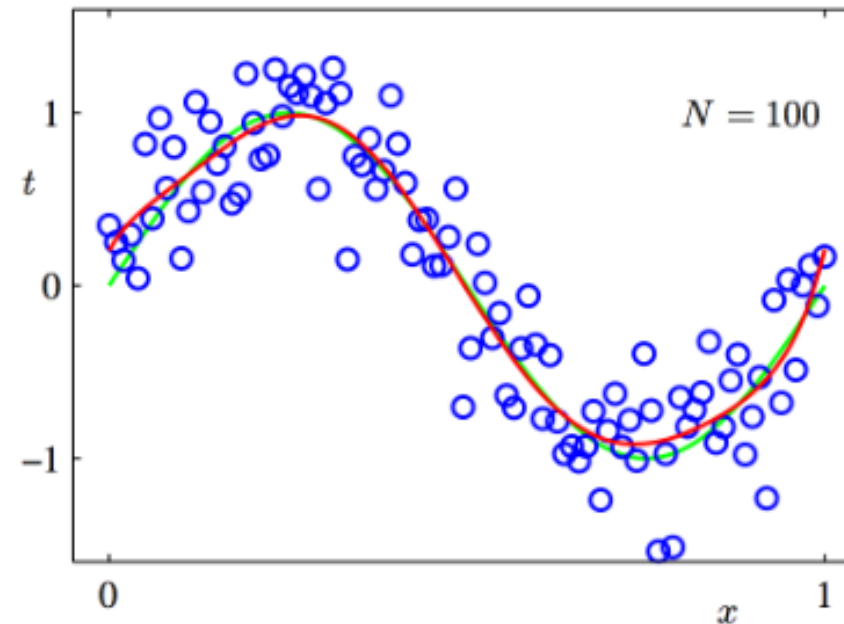
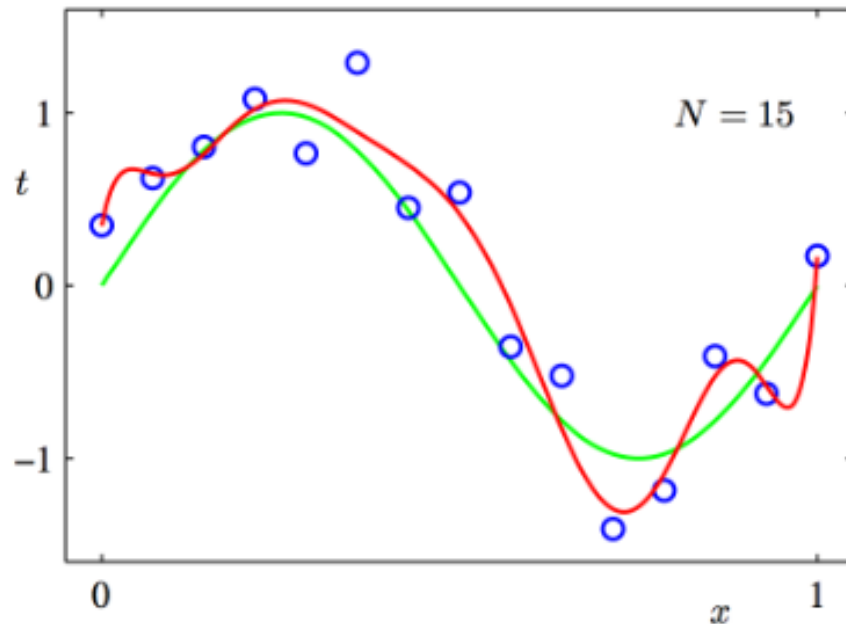
Generalization

- Generalization = model's ability to predict the held out data
- What is happening?
- Our model with $M = 9$ overfits the data (it models also noise)



Generalization

- Generalization = model's ability to predict the held out data
- What is happening?
- Our model with $M = 9$ overfits the data (it models also noise)
- Not a problem if we have lots of training examples



Outliers

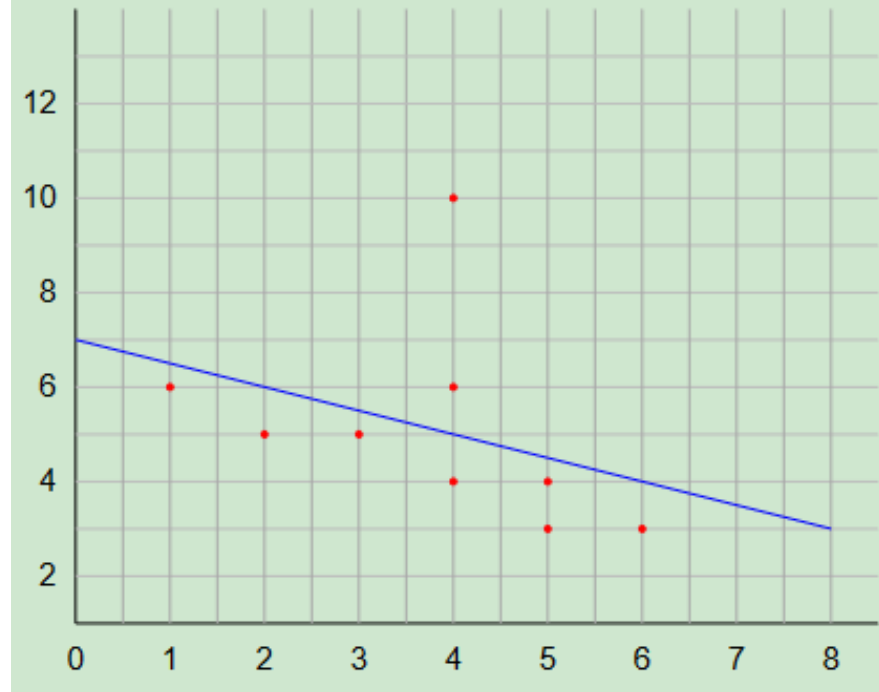
- Outlier
 - An outlier is a point with an unusually large residual (e.g., at least 2.5 standard deviations from the mean).
- Influential Point
 - An influential point is a point that exerts a inordinate influence on the regression line.
- An outlier may or may not be influential.
- An influential point may or may not be an outlier.

Example

- Consider the following data.

x	y
1	6
2	5
3	5
4	6
4	4
4	10
5	3
5	4
6	3

The scatterplot



Example

- The regression line is $\hat{y} = 7.0 - 0.5x$.

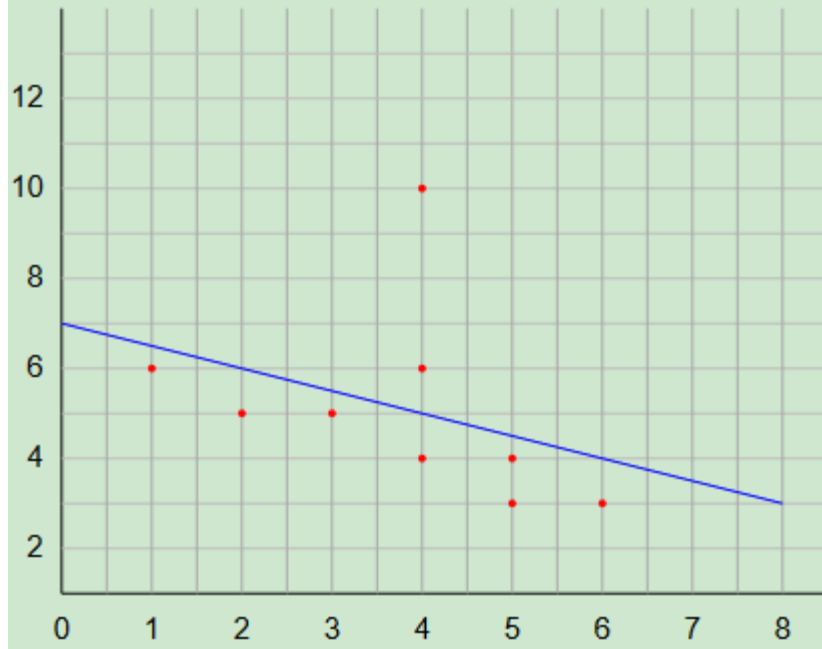
x	y	$\hat{y}$	$y - \hat{y}$
1	6	6.5	-0.5
2	5	6.0	-1.0
3	5	5.5	-0.5
4	6	5.0	1.0
4	4	5.0	-1.0
4	10	5.0	5.0
5	3	4.5	-1.5
5	4	4.5	-0.5
6	3	4.0	-1.0

- The mean residual is 0.0 (always) and the standard deviation of these residuals is 2.0.
- Thus, the residual 5.0 is 2.5 standard deviations above the mean, an outlier.
- But, is the point (4, 10) influential?
- Remove it and see what the effect is.

Example

Including the point (4, 10)

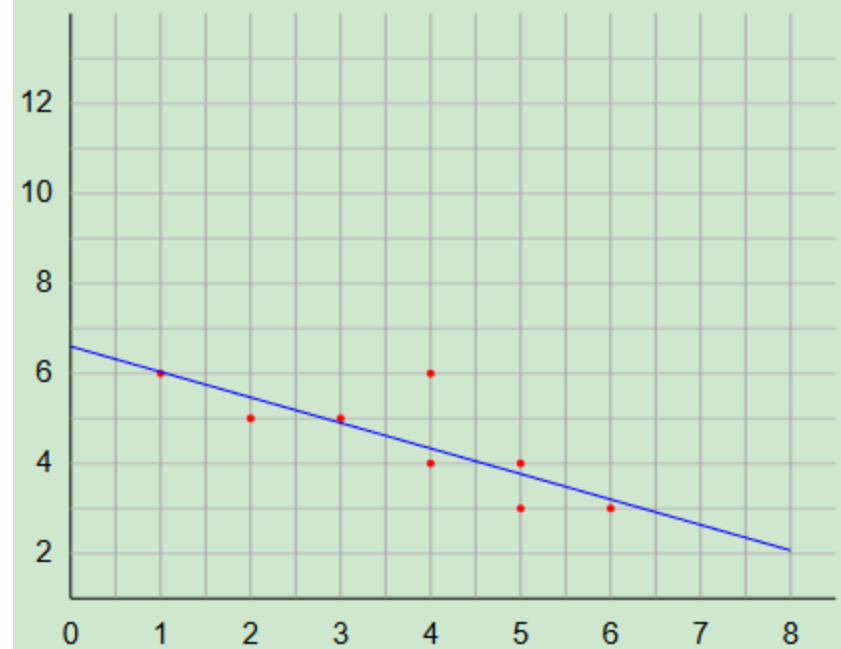
$$\hat{y} = 7.0 - 0.5x.$$



This is nearly the
same as →

Excluding the point (4, 10)

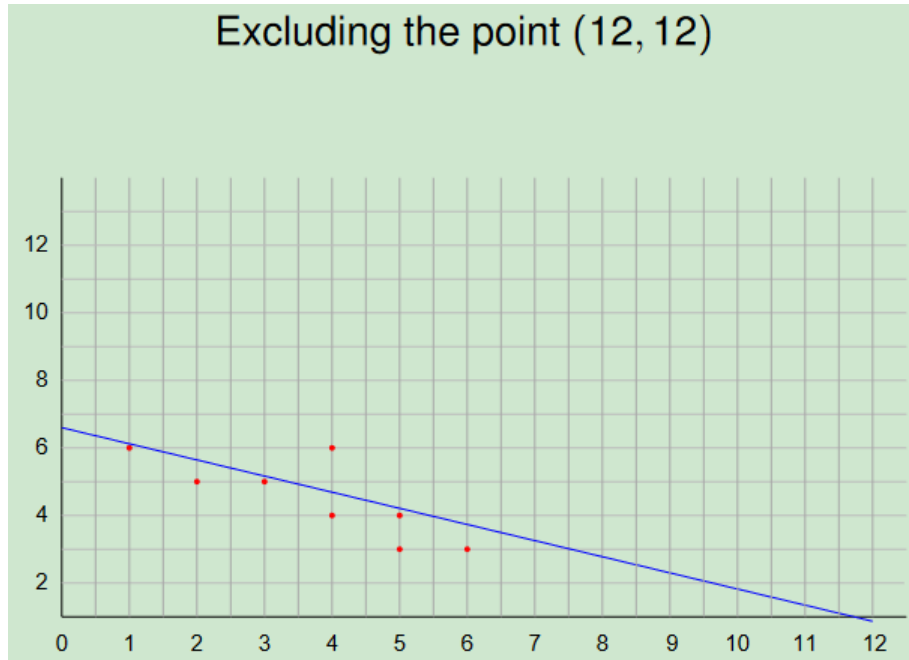
$$\hat{y} = 6.615 - 0.564x.$$



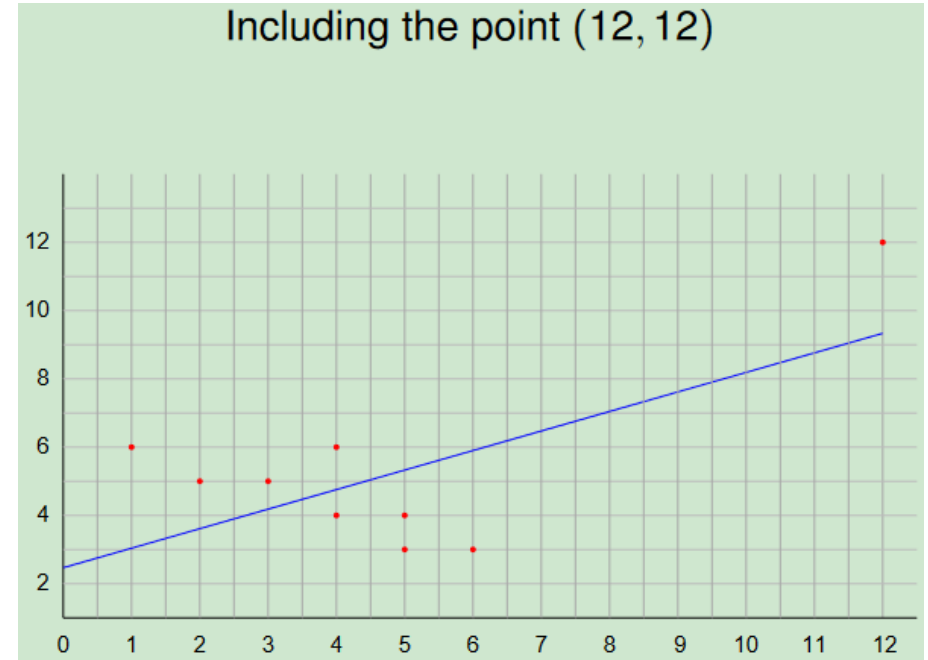
Example

- Now change the point (4, 10) to the point (12, 12).

x	y
1	6
2	5
3	5
4	6
4	4
5	3
5	4
6	3
12	12



$$\hat{y} = 6.615 - 0.564x$$



$$\hat{y} = 2.767 + 0.55x.$$