

# Big idea in data science

Using one algorithm (training algorithm) to train another (model algorithm) on data.

Legendre 1805 - orbits of comets, SSQ.

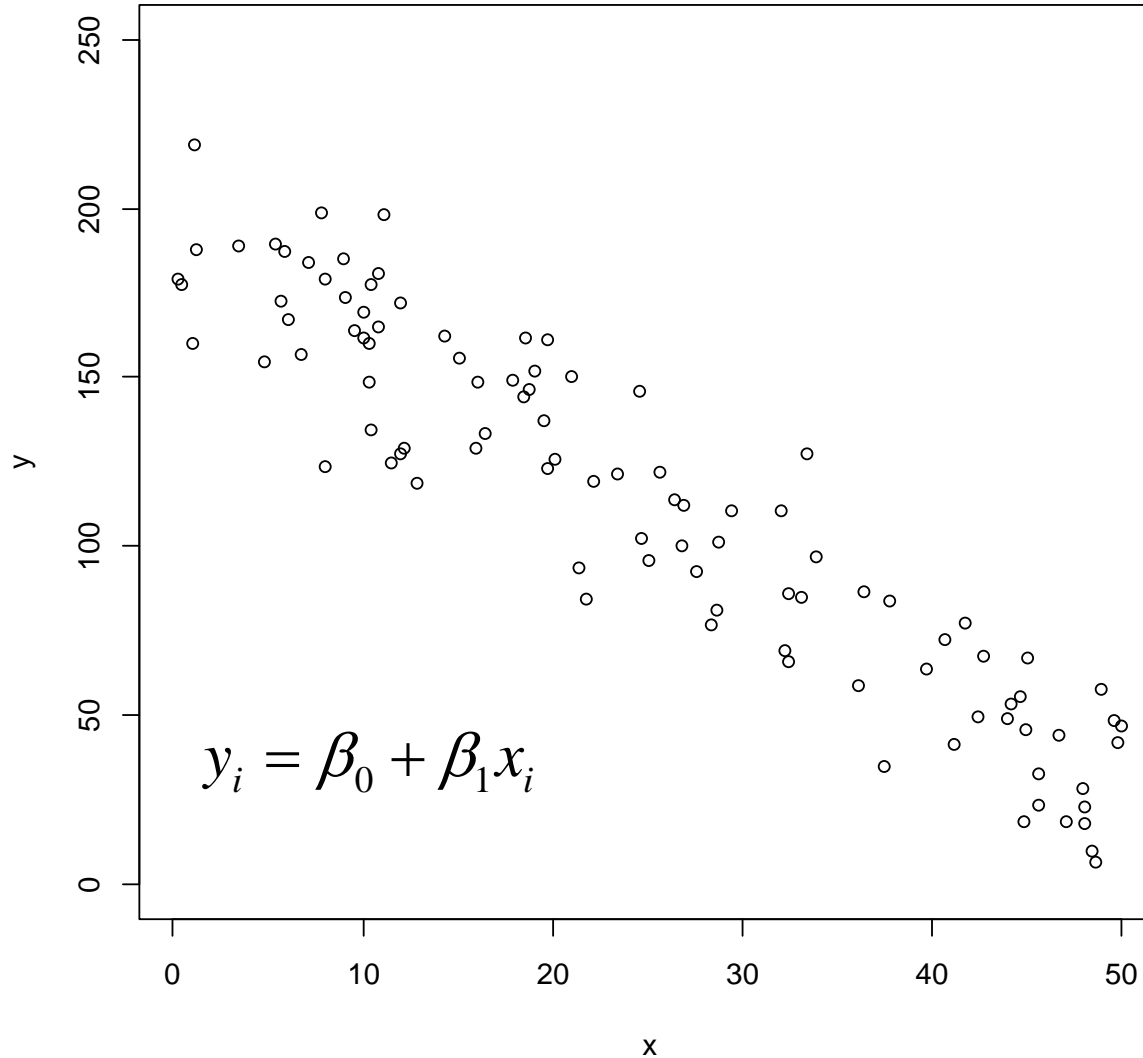
## Model algorithm



## Training algorithm

An algorithm to train a model algorithm on data  
syn. model fitting, calibration

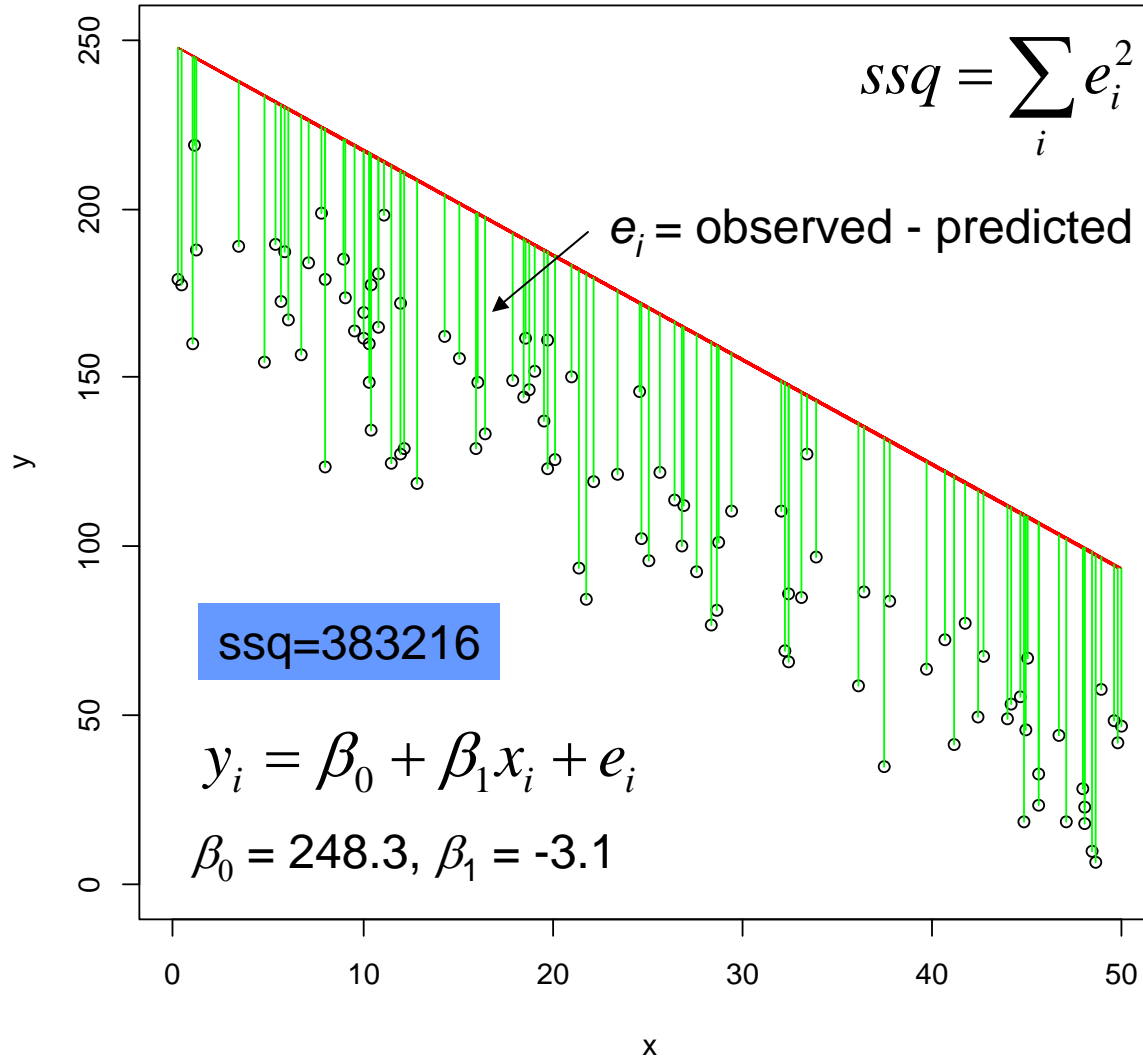
# Least squares algorithms



General algorithmic idea:

Vary model parameters until we find the parameter values that minimize the distance of the model from the data

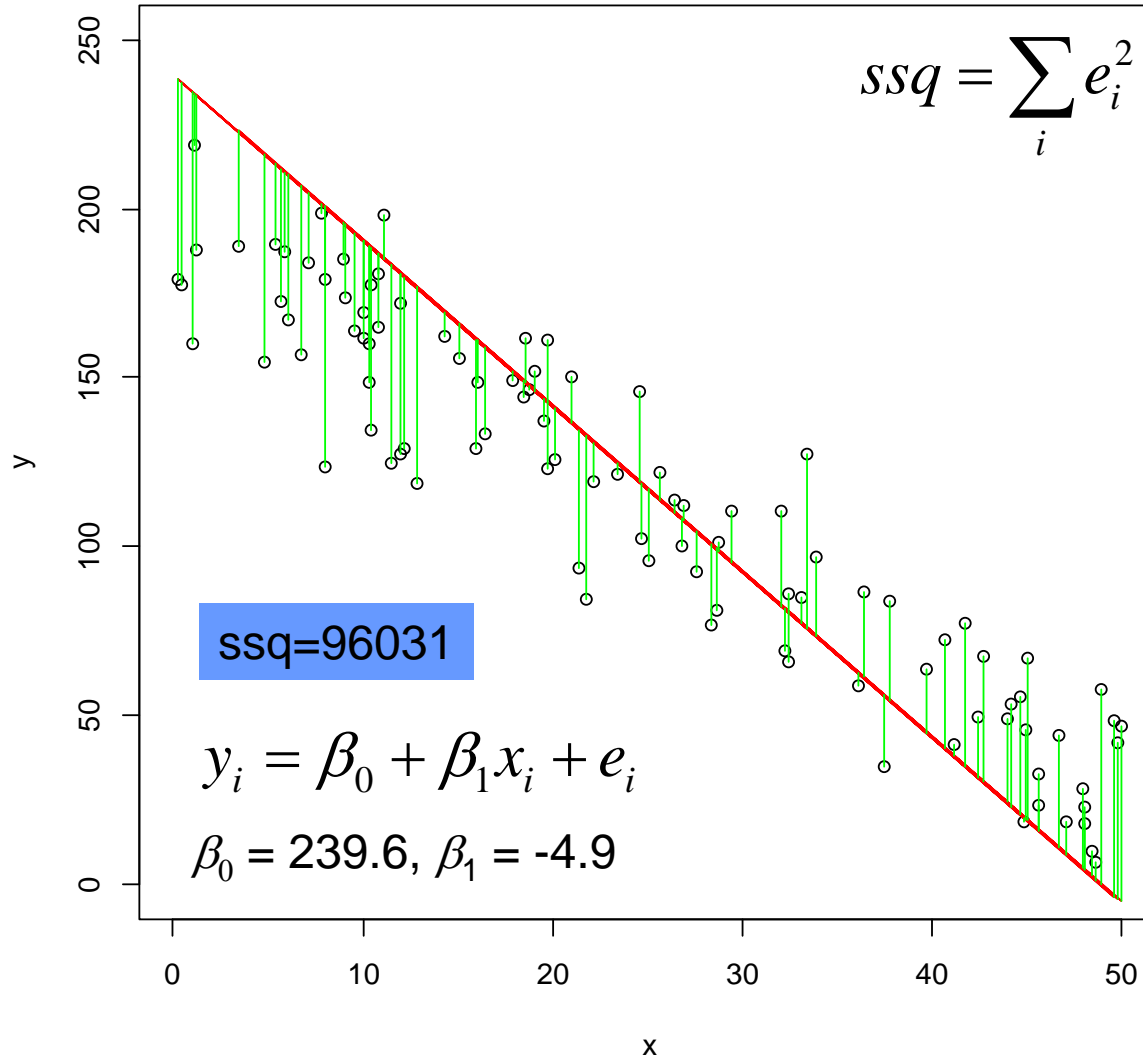
# Least squares algorithms



General algorithmic idea:

Vary model parameters until we find the parameter values that minimize the distance of the model from the data

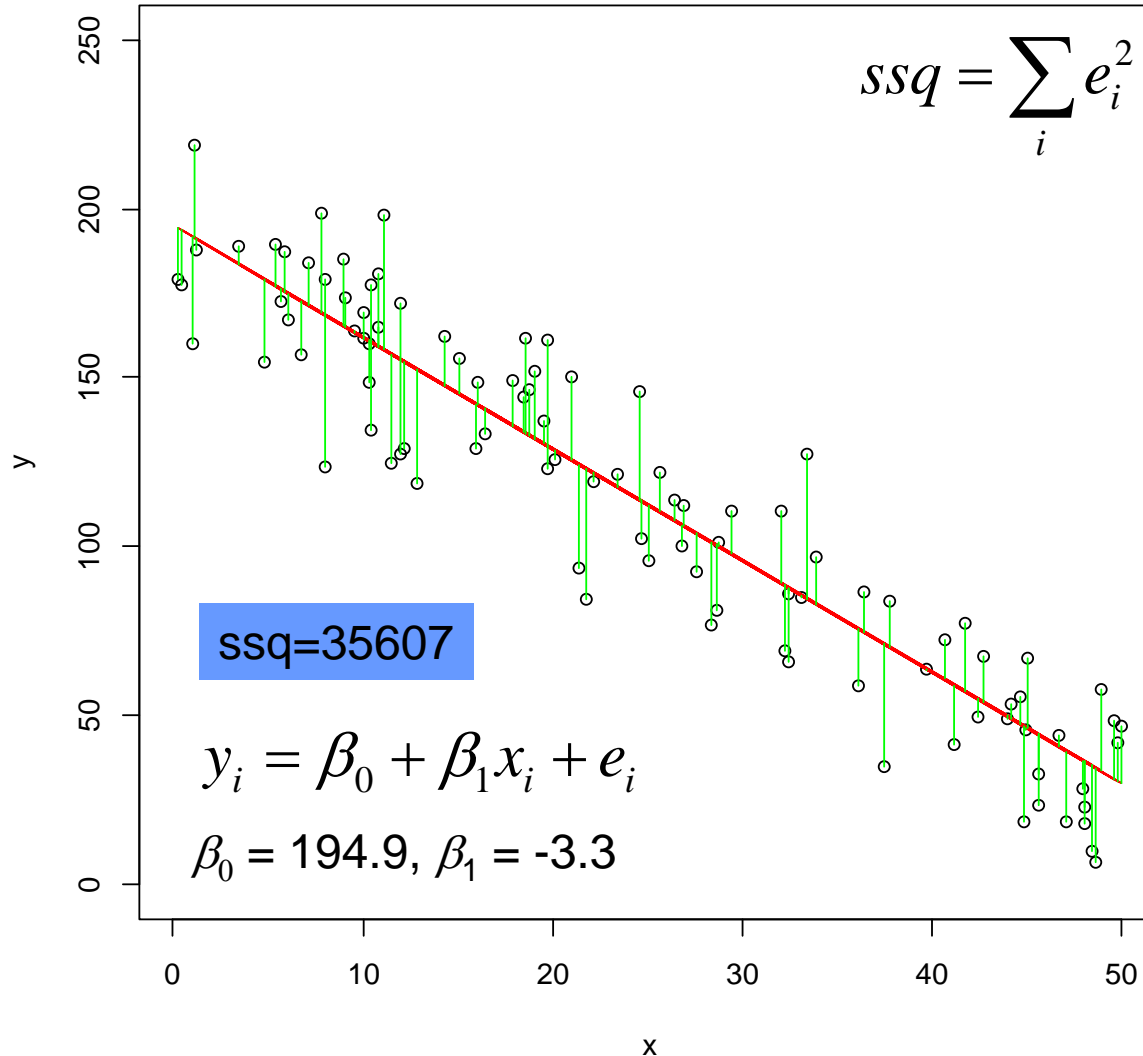
# Least squares algorithms



General algorithmic  
idea:

Vary model  
parameters until we  
find the parameter  
values that minimize  
the distance of the  
model from the data

# Least squares algorithms



General algorithmic idea:

Vary model parameters until we find the parameter values that minimize the distance of the model from the data

# Optimization algorithms

- Systematically try all combinations for  $\beta_0$  and  $\beta_1$  - Grid search algorithms
- Narrowing in - keep changing parameters in the direction that leads to lower SSQ - Descent algorithms
- Try random values for  $\beta_0$  and  $\beta_1$  - Monte Carlo algorithms
- Solve for parameters using math - Analytical or numerical algorithms

# Developing an algorithm: an example

Key points:

Pseudocode

3 Phases

Top down refinement

# Grid search algorithm

## Pseudocode

For each value of  $\beta_0$   
    For each value of  $\beta_1$   
        Calculate sum of squares



# Grid search algorithm

## Pseudocode

Read in data

Set up values of  $\beta_0$  and  $\beta_1$  to try

Set up storage for ssq,  $\beta_0$ ,  $\beta_1$

For each value of  $\beta_0$

For each value of  $\beta_1$

Calculate sum of squares

Store ssq,  $\beta_0$ ,  $\beta_1$

Plot sum of squares profiles (ssq vs  $\beta_0$ , ssq vs  $\beta_1$ )

Report best ssq,  $\beta_0$ ,  $\beta_1$

Plot fitted model with the data

# Grid search algorithm

## Pseudocode

Read in data

Set up values of  $\beta_0$  and  $\beta_1$  to try

Set up storage for ssq,  $\beta_0$ ,  $\beta_1$

For each value of  $\beta_0$

    For each value of  $\beta_1$

        Calculate sum of squares

        Store ssq,  $\beta_0$ ,  $\beta_1$

Plot sum of squares profiles (ssq vs  $\beta_0$ , ssq vs  $\beta_1$ )

Report best ssq,  $\beta_0$ ,  $\beta_1$

Plot fitted model with the data

Initialization  
phase

Calculation  
phase

Termination  
phase

# Grid search algorithm

## Pseudocode

Read in data

Set up values of  $\beta_0$  and  $\beta_1$  to try

Set up storage for ssq,  $\beta_0$ ,  $\beta_1$

For each value of  $\beta_0$

    For each value of  $\beta_1$

        Calculate sum of squares

        Store ssq,  $\beta_0$ ,  $\beta_1$

Plot sum of squares profiles (ssq vs  $\beta_0$ , ssq vs  $\beta_1$ )

Report best ssq,  $\beta_0$ ,  $\beta_1$

Plot fitted model with the data

Top down  
refinement

# Grid search algorithm

## Pseudocode

Read in data

Set up values of  $\beta_0$  and  $\beta_1$  to try

Set up storage for ssq,  $\beta_0$ ,  $\beta_1$

For each value of  $\beta_0$

    For each value of  $\beta_1$

        Calculate model predictions

        Calculate deviations

        Sum squared deviations

        Store ssq,  $\beta_0$ ,  $\beta_1$

Plot sum of squares profiles (ssq vs  $\beta_0$ , ssq vs  $\beta_1$ )

Report best ssq,  $\beta_0$ ,  $\beta_1$

Plot fitted model with the data

Top down  
refinement