

Algorithms in data science

- Model algorithm
- Training algorithm
- Inference algorithm

Algorithms in data science

- Model algorithm
 - Often explicitly expressed as equations
 - Sometimes a set of rules
 - Has parameters
- Training algorithm
- Inference algorithm

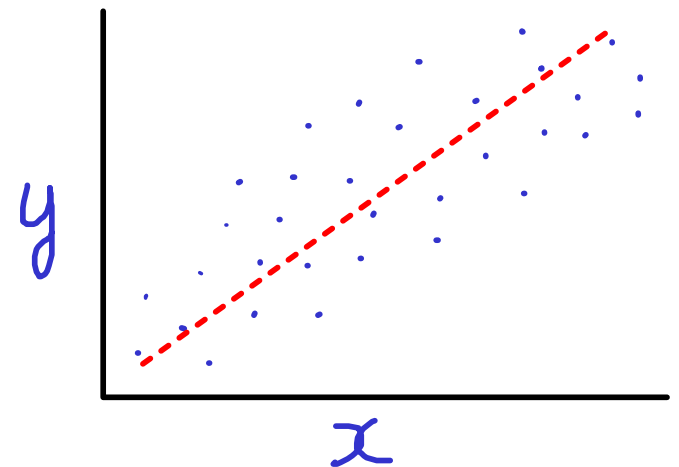
How can a model be an algorithm?

Model

$$y_i = \beta_0 + \beta_1 x_i$$

Data table

i	y	x
1	28.4	10.2
2	47.6	15.7
...		
...		
85	35.1	12.9



How can a model be an algorithm?

Model

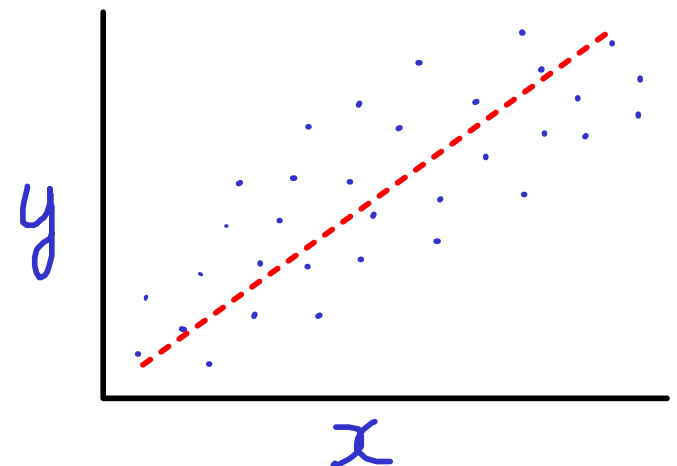
$$y_i = \beta_0 + \beta_1 x_i$$

Algorithm (this version is atomic code)

```
for ( i in 1:n ) {  
    y[i] = b_0 + b_1 * x[i]  
}
```

Data table

i	y	x
1	28.4	10.2
2	47.6	15.7
...		
...		
85	35.1	12.9



How can a model be an algorithm?

Model

$$y_i = \beta_0 + \beta_1 x_i$$

Algorithm (this version is atomic code)

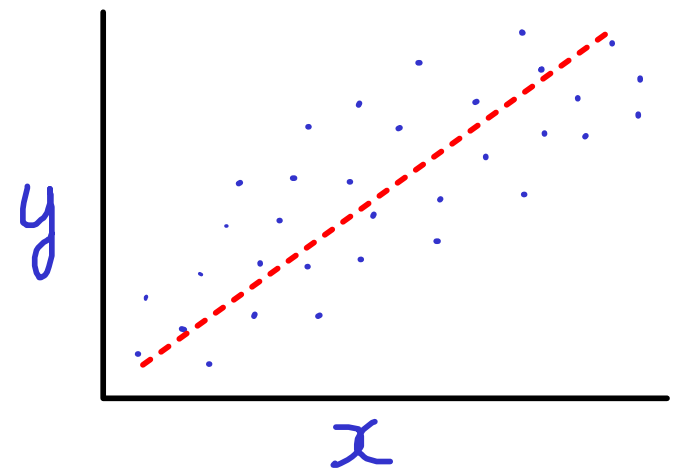
```
for ( i in 1:n ) {  
  y[i] = b_0 + b_1 * x[i]  
}
```

Vectorized R code

```
y <- b_0 + b_1 * x
```

Data table

i	y	x
1	28.4	10.2
2	47.6	15.7
...		
...		
85	35.1	12.9



Algorithms in data science

- Model algorithm
- Training algorithm
 - An algorithm to train a model algorithm on data
 - syn. model fitting, calibration
 - e.g. least squares, maximum likelihood
 - try typing `lm` into the R console
- Inference algorithm

Algorithms in data science

- Model algorithm
- Training algorithm
- Inference algorithm
 - **looking back**: considering all the ways data could have happened
 - **looking forward**: predicting new data and testing against them

Workflow algorithms in DS

- Pipelines, standard conventions ...

Master Data Science Algorithm

Plan for data

Acquire data

Manage data

Analyze data

Infer from data

Report about data