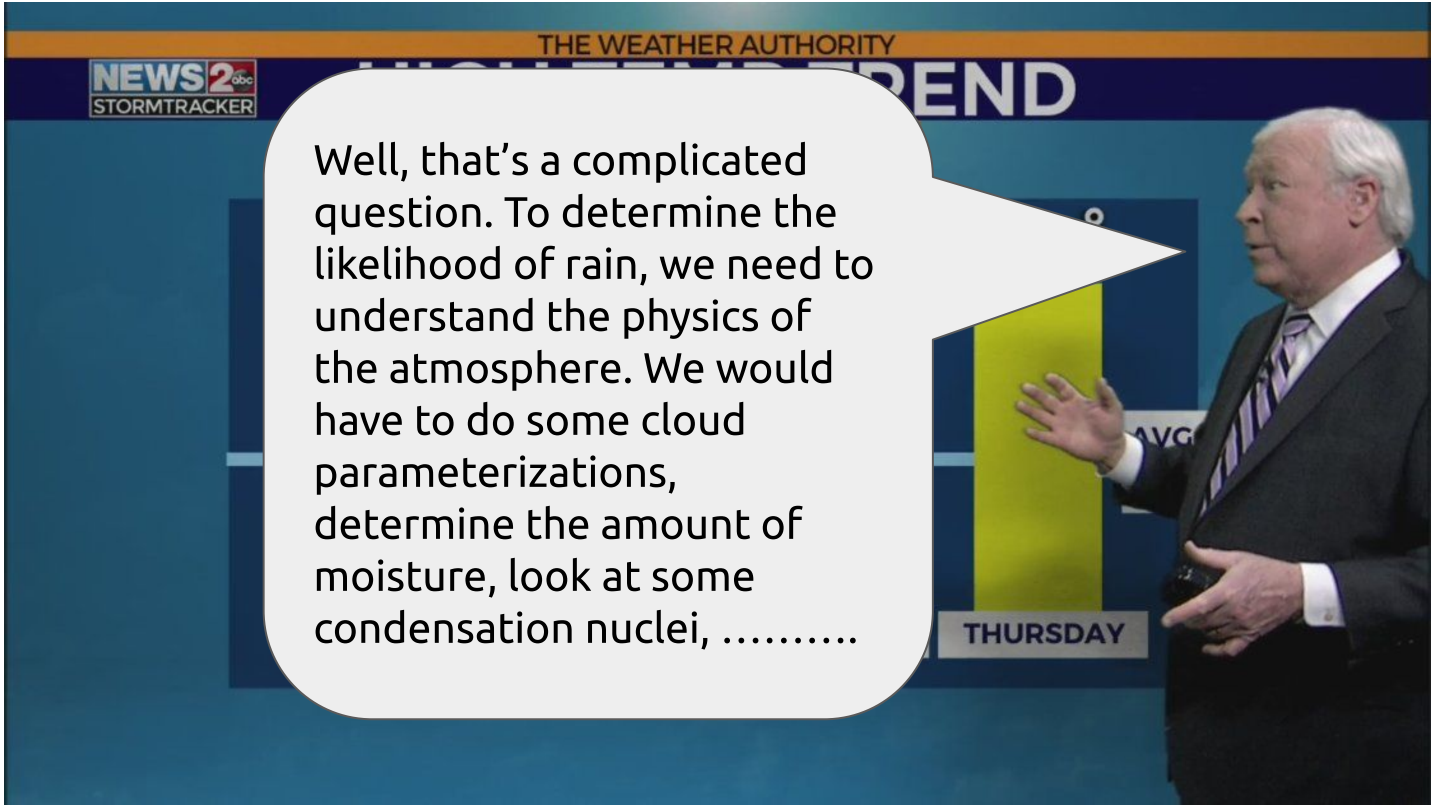
Introduction to Supervised Learning

Question - Is it going to rain today?

Question - Is it going to rain today?



Question - Is it going to rain today?

You look outside and it looks like this:

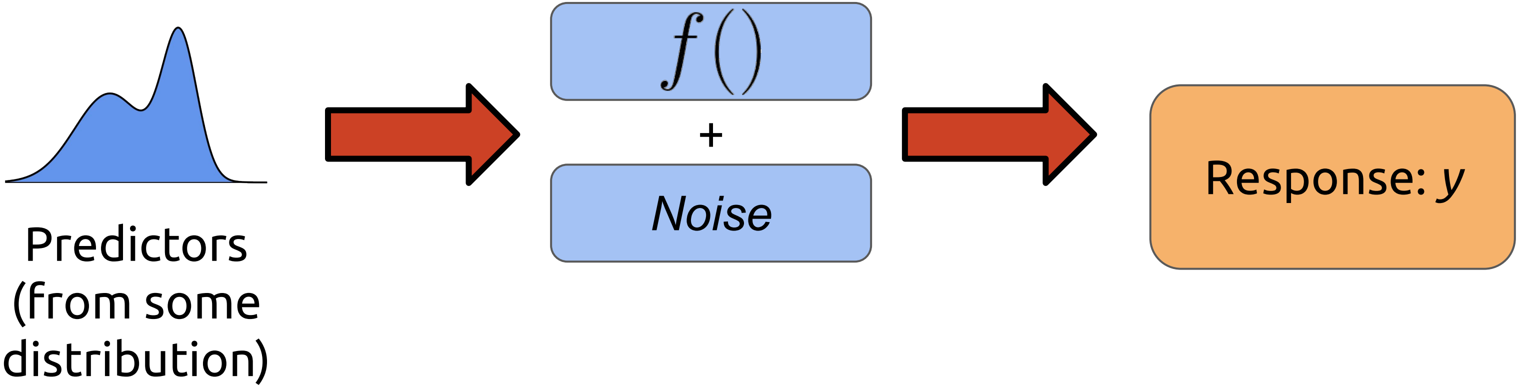
Question - Is it going to rain today?

You look outside and it looks like this:

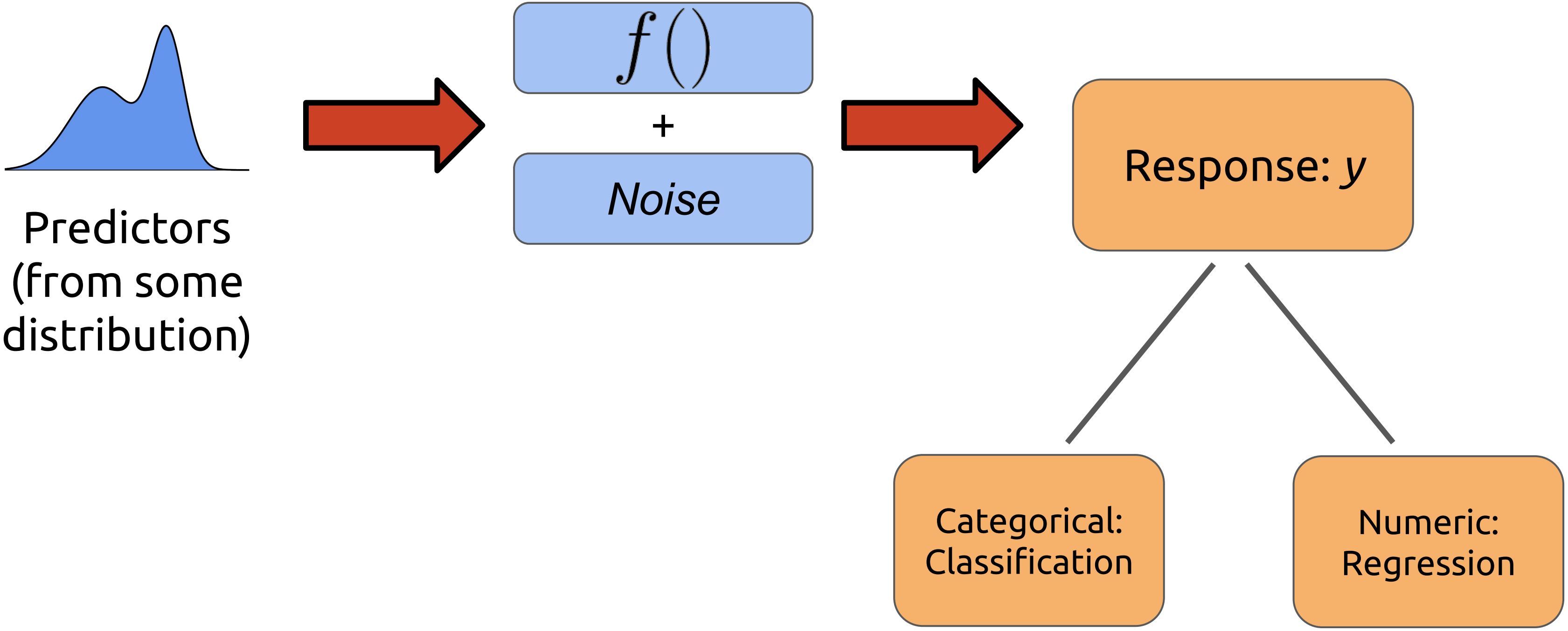
Question - Is it going to rain today?

You look outside and it looks like this:

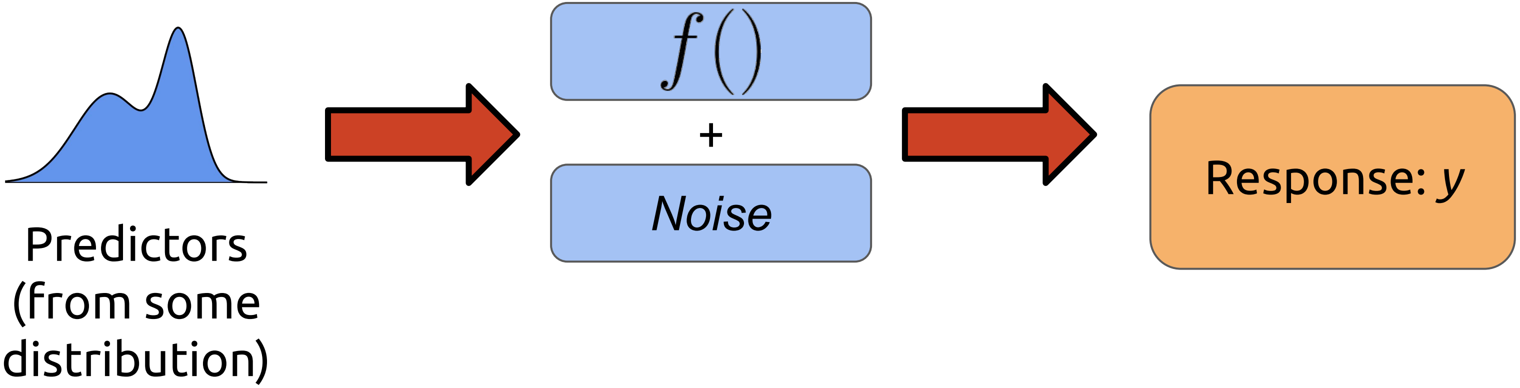
Supervised Learning - Setup



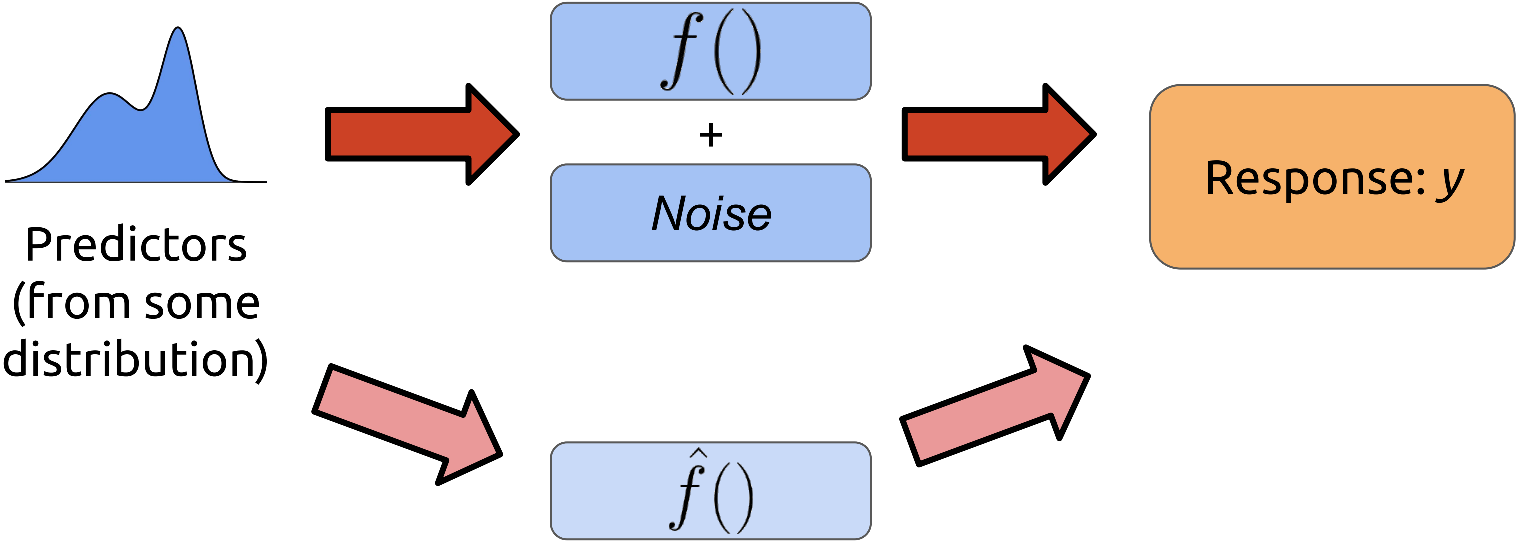
Supervised Learning - Setup



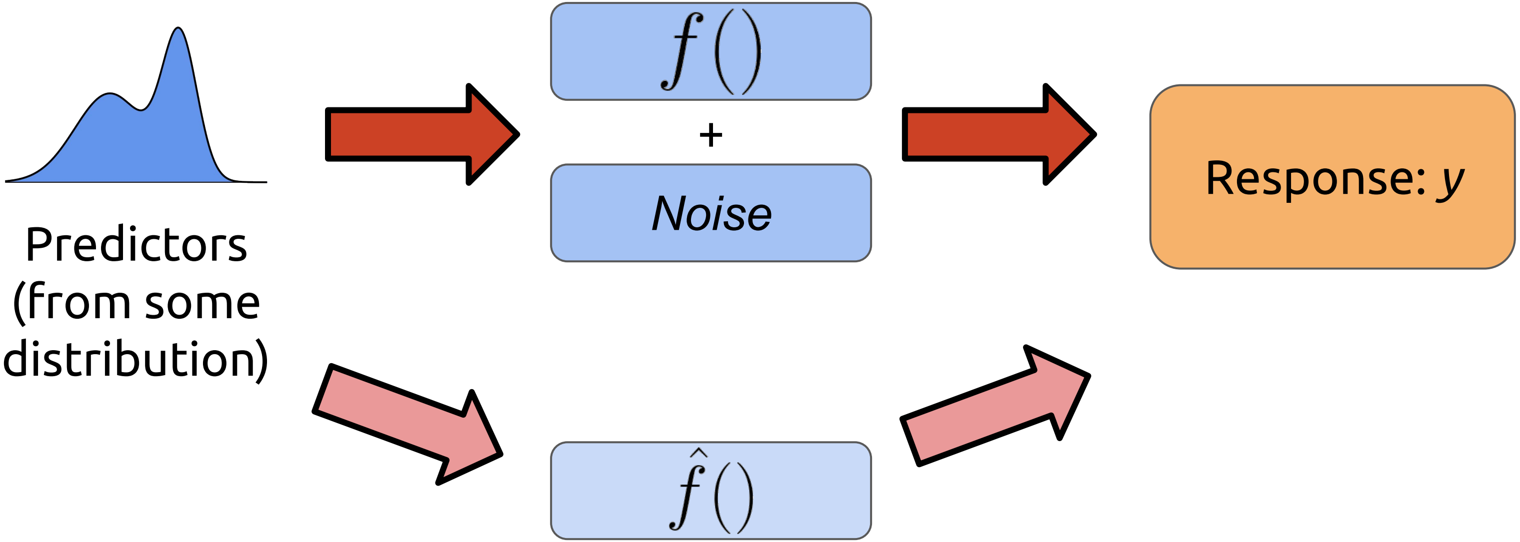
Supervised Learning - Setup



Supervised Learning - Goals

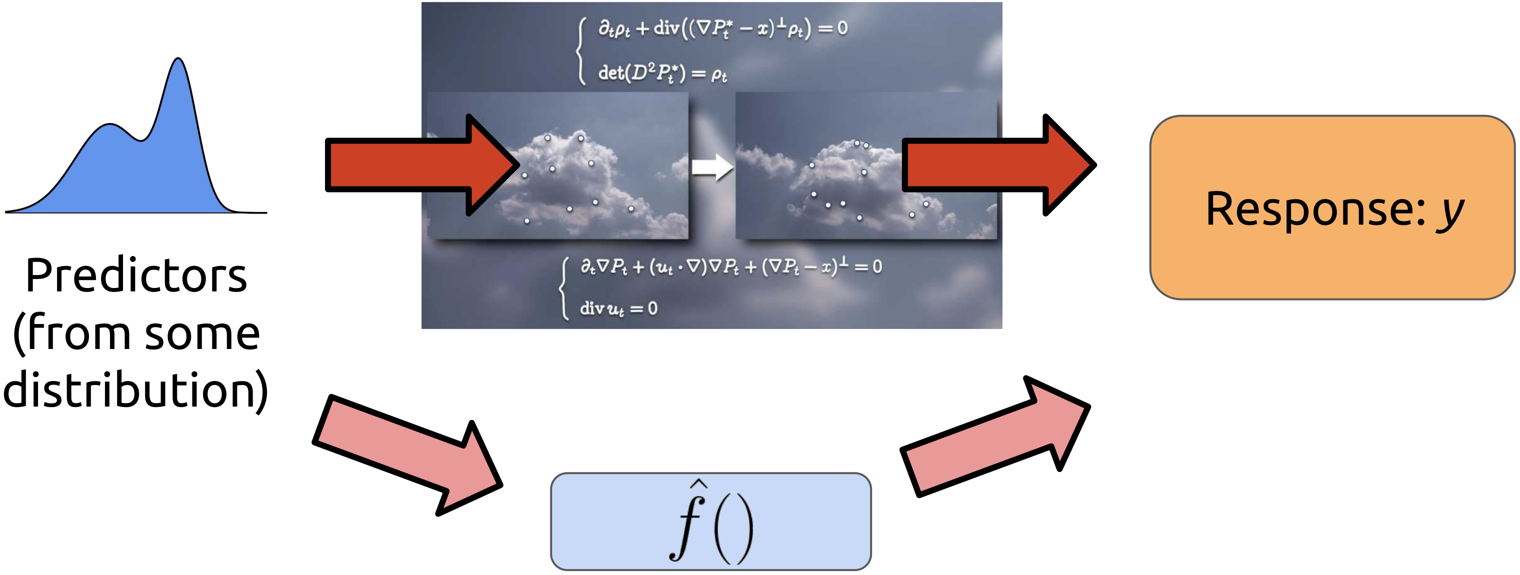


Supervised Learning - Goals

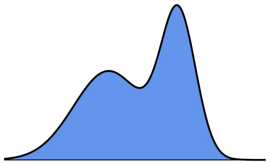


Goal: Choose a function so that the our predictions are close (on average) to the true values.

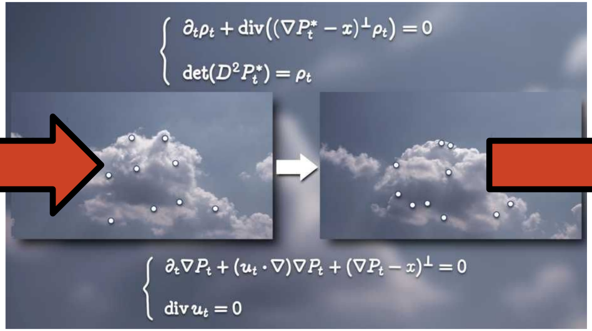
Supervised Learning - Goals



Supervised Learning - Goals



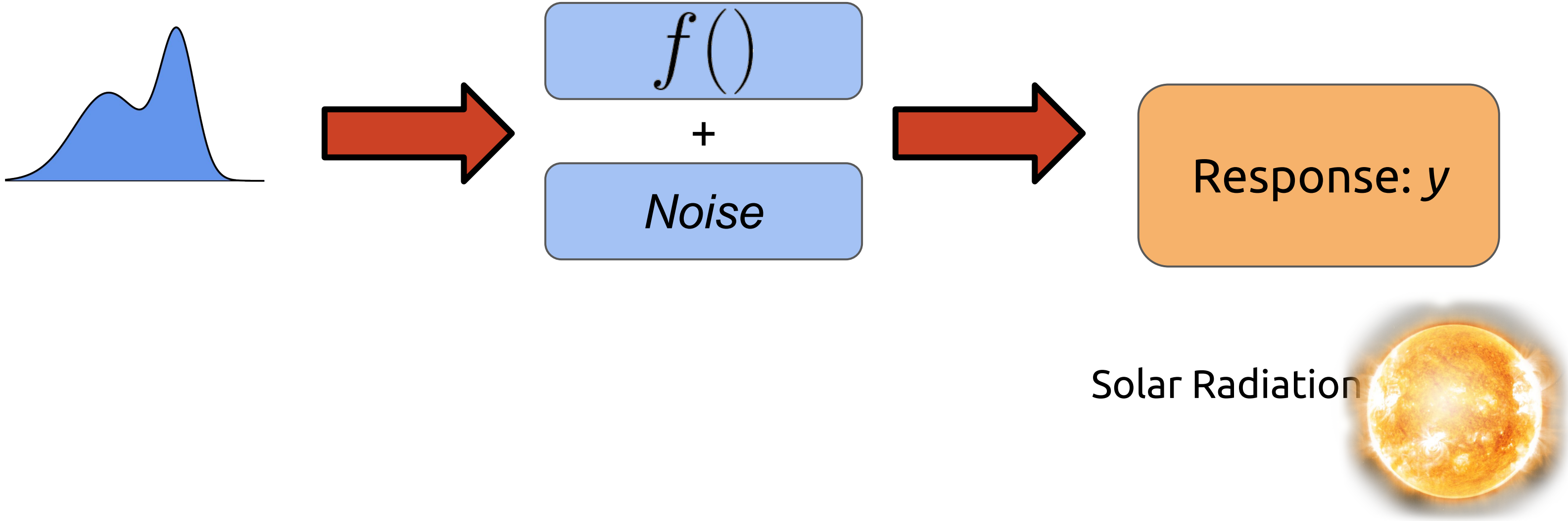
Predictors (from some distribution)



Response: *y*

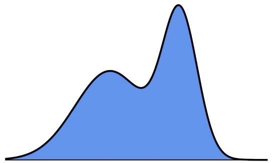
My knee is acting up. Must be rain coming.

Example - Weather Prediction



https://www.ai.uga.edu/sites/default/files/inline-files/theses/sanders_william_s_201712_ms.pdf

Example - Weather Prediction



air temperature
precipitation rate
visibility
wind speed
wind direction
dew point temperature
air pressure
relative humidity

$f()$

+

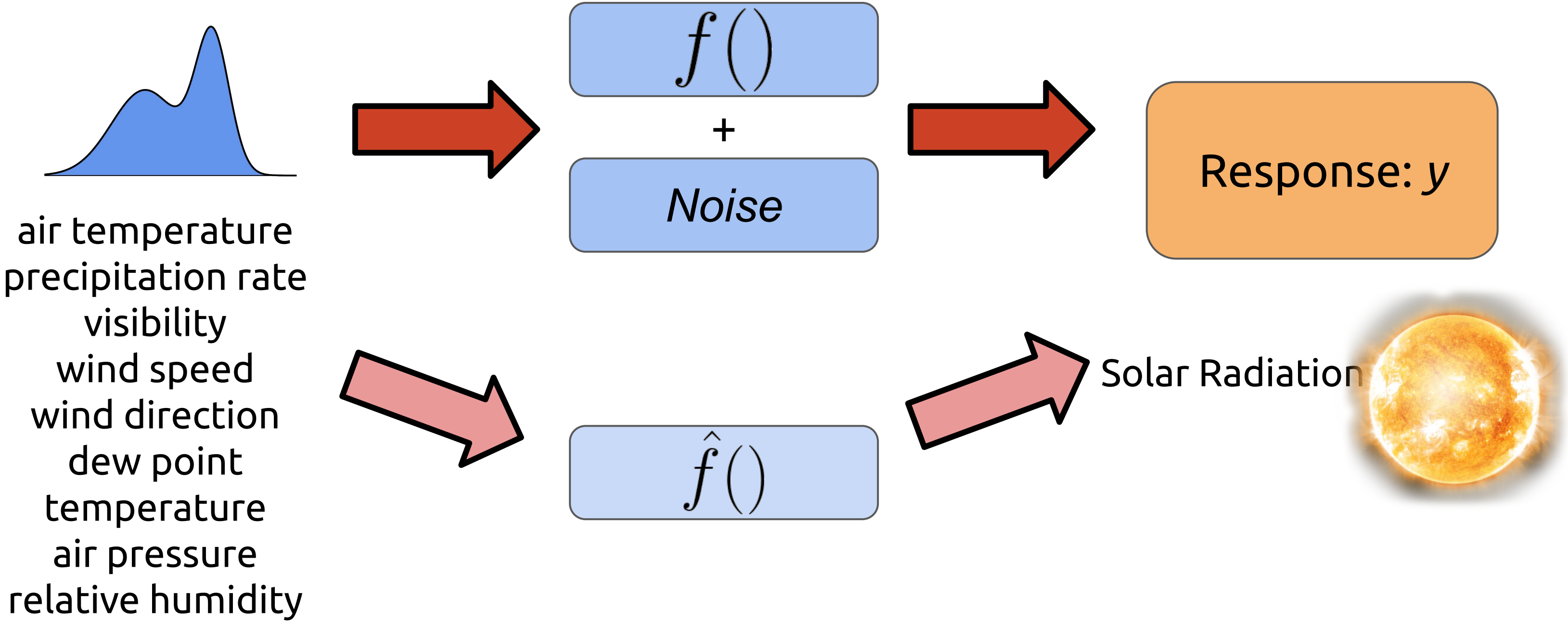
Noise

Response: *y*

Solar Radiation

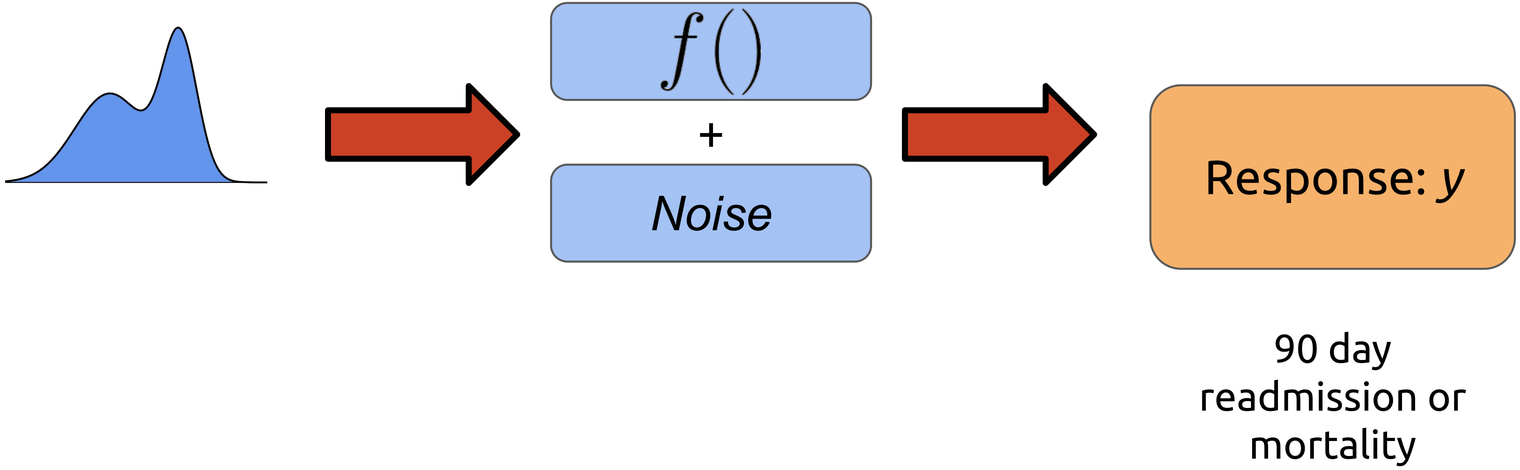
https://www.ai.uga.edu/sites/default/files/inline-files/theses/sanders_william_s_201712_ms.pdf

Example - Weather Prediction



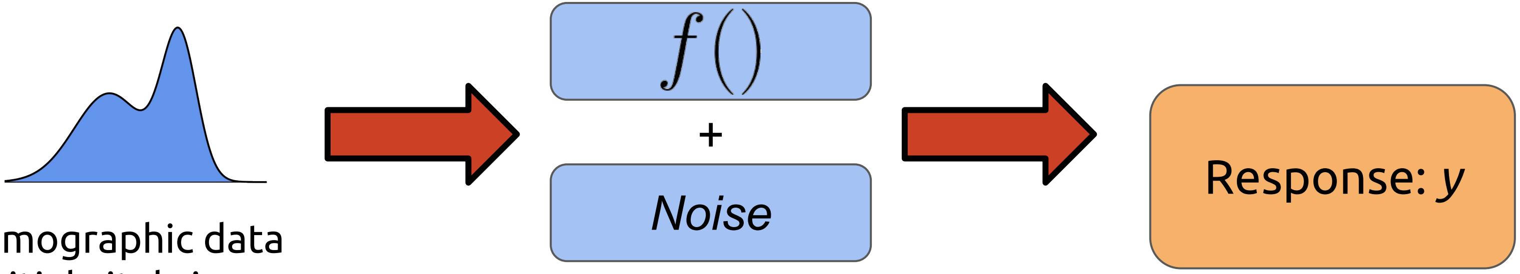
https://www.ai.uga.edu/sites/default/files/inline-files/theses/sanders_william_s_201712_ms.pdf

Example - Readmission or Death of Stroke Patients



https://www.mdpi.com/2076-3417/10/18/6337/pdf

Example - Readmission or Death of Stroke Patients

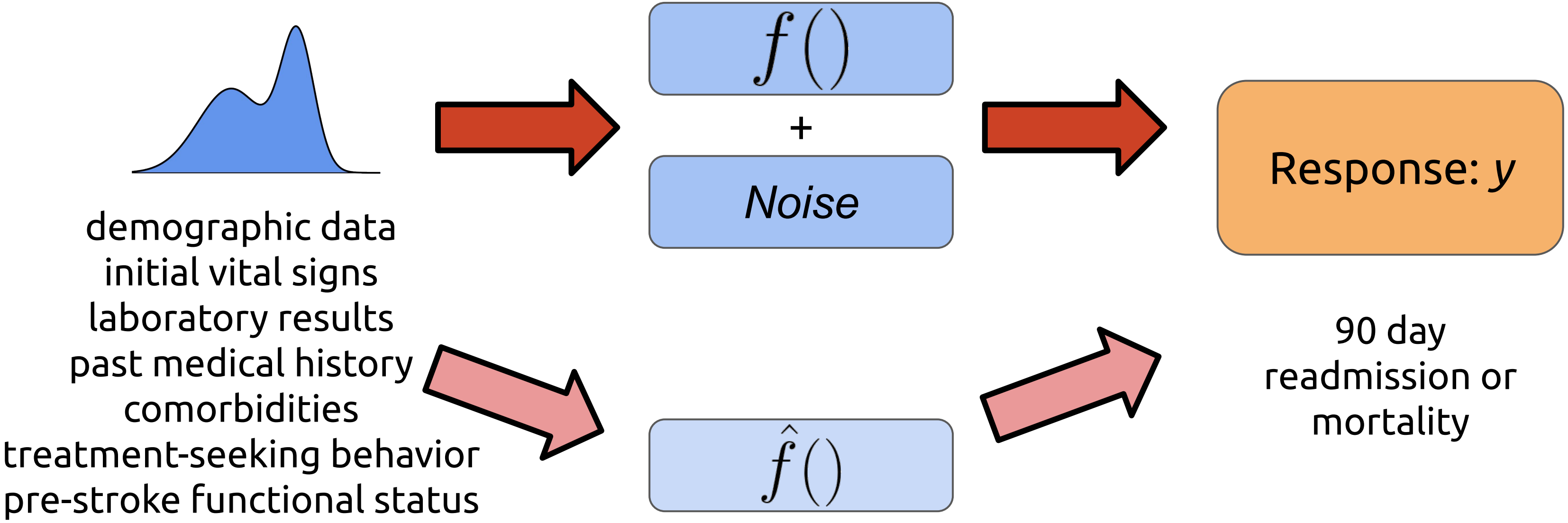


demographic data
initial vital signs
laboratory results
past medical history
comorbidities
treatment-seeking behavior
pre-stroke functional status

90 day
readmission or
mortality

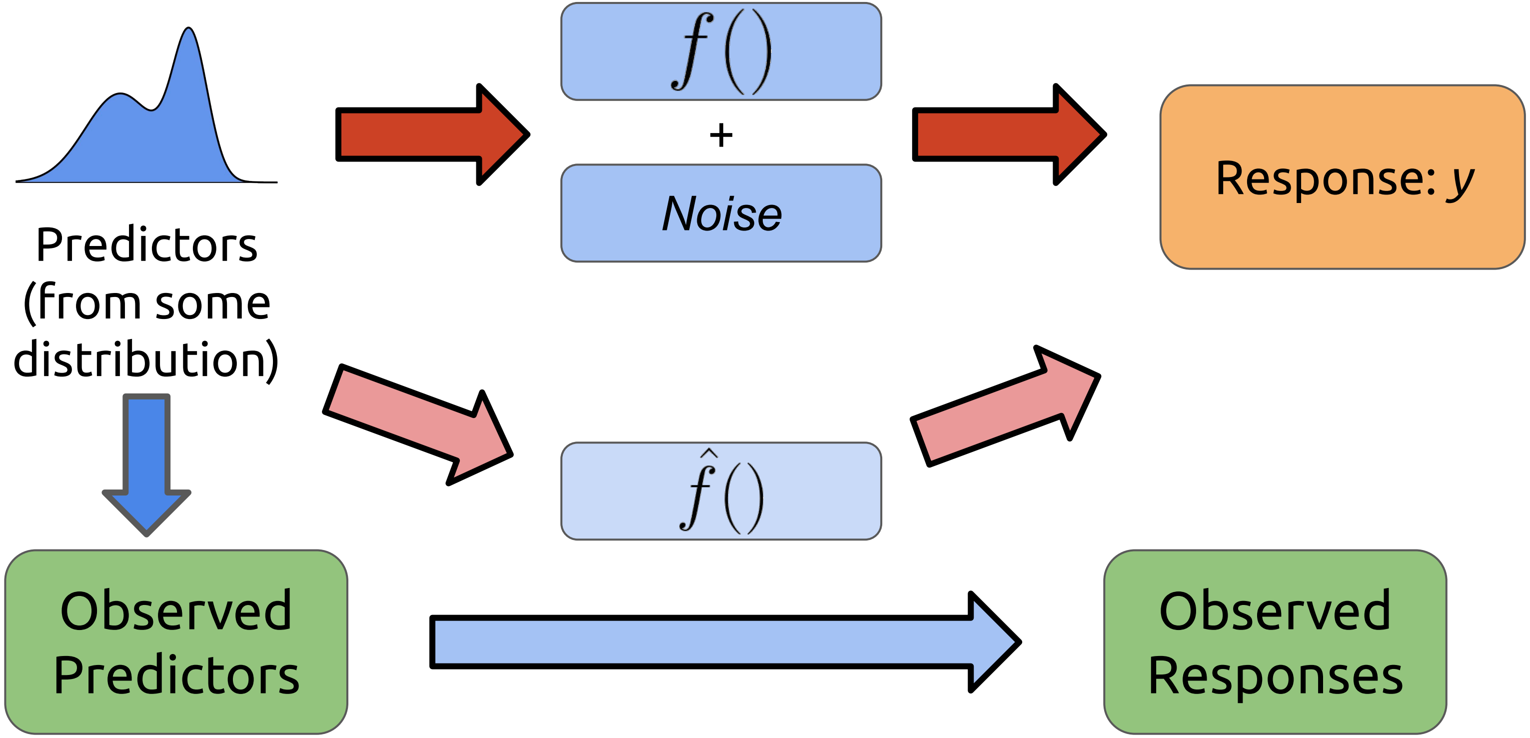
https://www.mdpi.com/2076-3417/10/18/6337/pdf

Example - Readmission or Death of Stroke Patients



https://www.mdpi.com/2076-3417/10/18/6337/pdf

Supervised Learning - How



Supervised Learning - Goals

To measure how “good” our model is, we need some way to measure “error” (eg. mean squared error).

Our goal is to minimize the expected loss over *new* data.

Important: We are not trying to minimize loss over the observed data (which is often very easy to do), but to minimize the *generalization error* - the performance on unseen data.

Supervised Learning - How?

We need to pick a way to make predictions from our available training data.

Supervised Learning - How?

We need to pick a way to make predictions from our available training data.

There are many, many ways to do this.

Supervised Learning - How?

We need to pick a way to make predictions from our available training data.

There are many, many ways to do this.

For example, we can pick a functional form for $\hat{f}()$

Linear Regression

Given k predictors $x^{(1)}$, $x^{(2)}$,...,$x^{(k)}$, linear regression uses the following equation to predict the target variable:

$$\hat{f}(\vec{x}) = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \cdots + \beta_k x^{(k)}$$

Here, β_0, β_1,...,β_k are constants that are determined by using the available training data.

Linear Regression

Example: We might want to try and predict home price (our target) based on square footage (sqft), number of bedrooms (br), and number of floors (floors).

The model we will use to make predictions will look like:

$$\hat{f}(\vec{x}) = \beta_0 + \beta_1 \cdot (\text{sqft}) + \beta_2 \cdot (\text{br}) + \beta_3 \cdot (\text{floors})$$

Linear Regression

Example: We might want to try and predict home price (our target) based on square footage (sqft), number of bedrooms (br), and number of floors (floors).

The model we will use to make predictions will look like:

$$\hat{f}(x) = 40000 + 180 \cdot (\text{sqft}) + 15000 \cdot (\text{br}) + 30000 \cdot (\text{floors})$$

Linear Regression

How do we find the values for the coefficients?

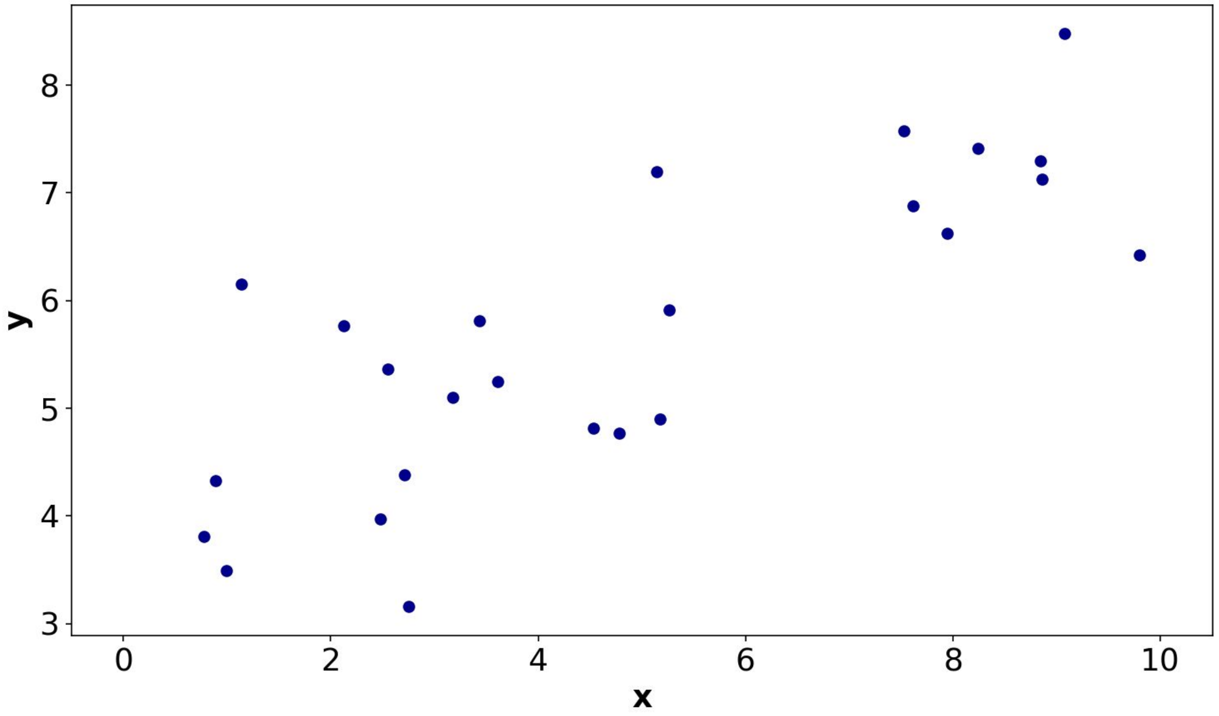
Linear Regression

How do we find the values for the coefficients?

The usual way to do it is to minimize the total squared residuals between the predicted and actual values for the data used to fit/train the model.

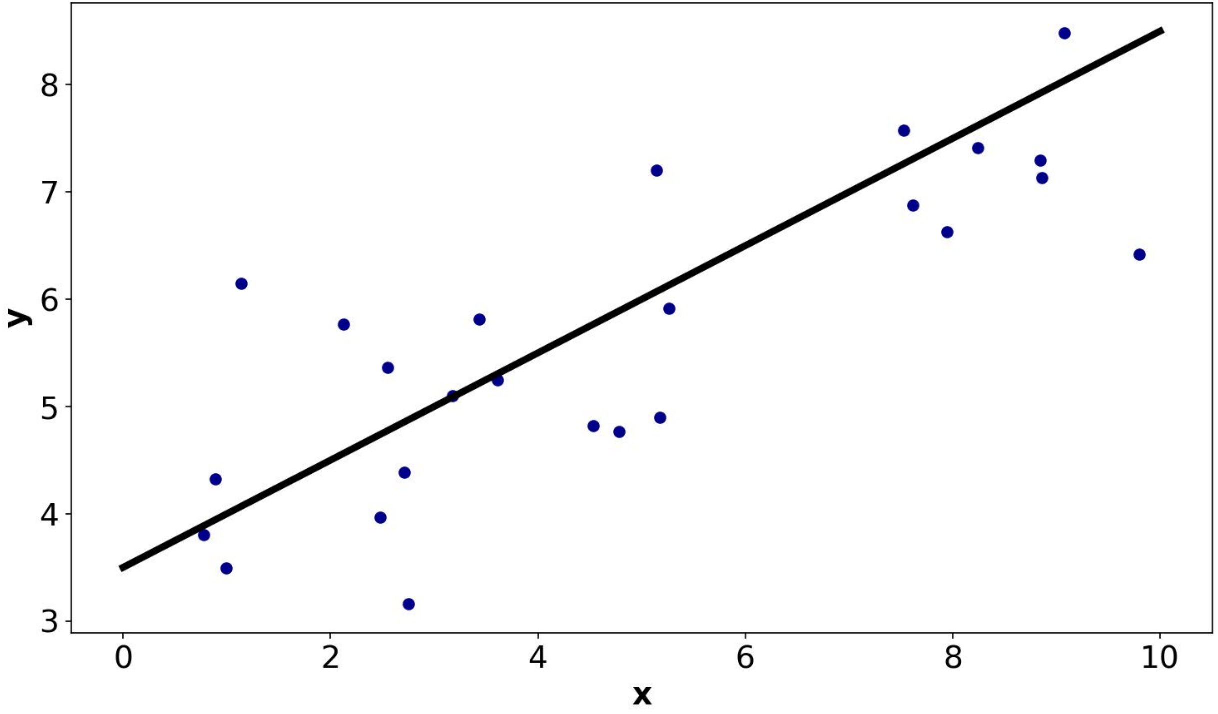
$$RSS = \sum_{i=1}^{n} (y_i - \hat{f}(\vec{x}_i))^2$$

Linear Regression



Example: Let's say we have this data available. We want to predict *y* based on our one predictor, *x*.

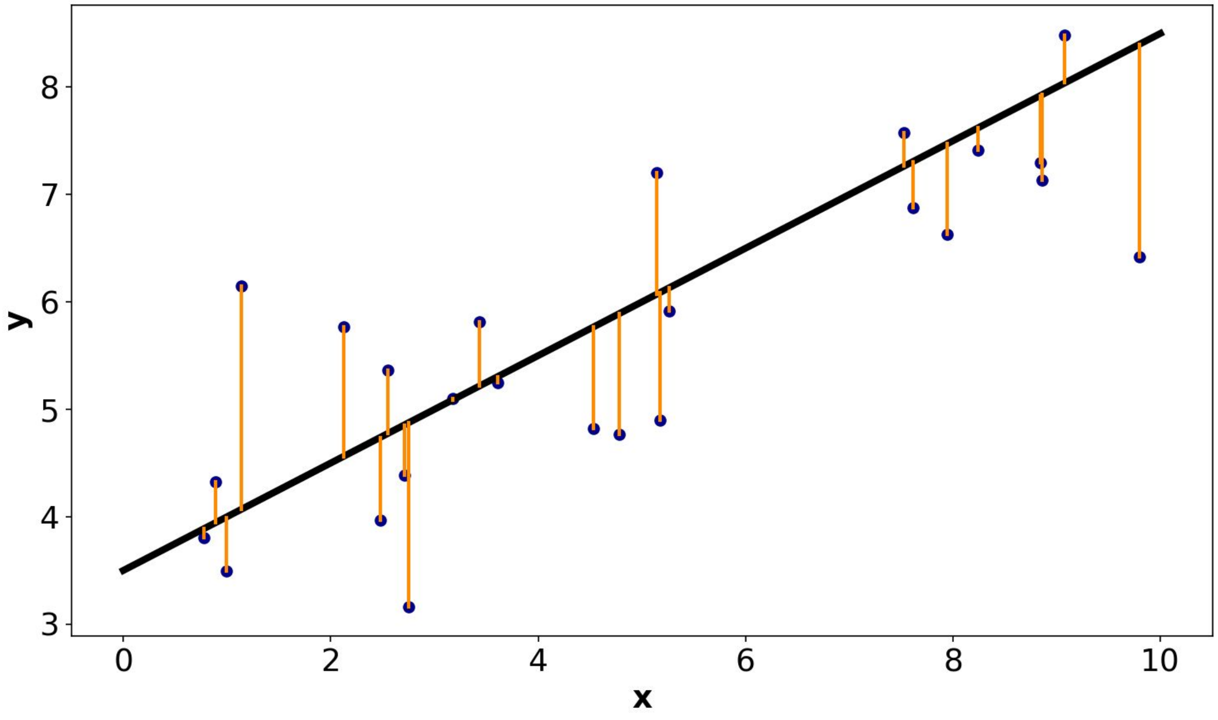
Linear Regression



One possible line:

$y = 3.5 + 0.5x$

Linear Regression



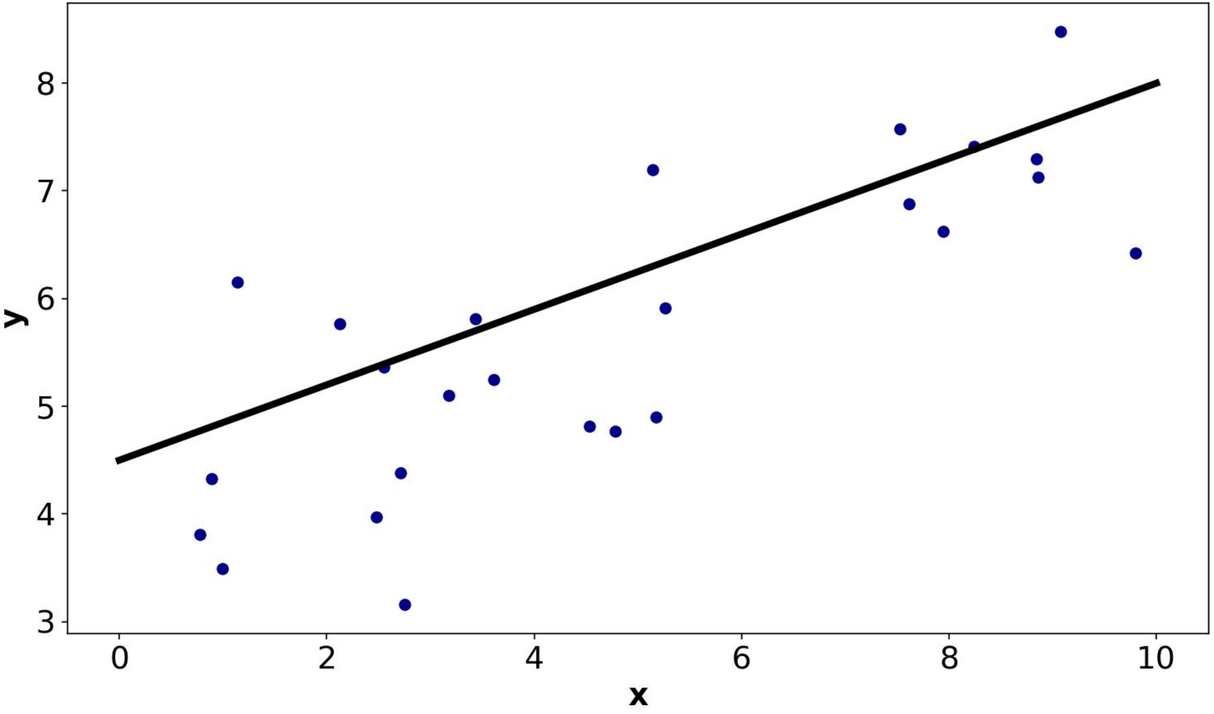
One possible line:

$y = 3.5 + 0.5x$

For this line,

RSS = 20.36

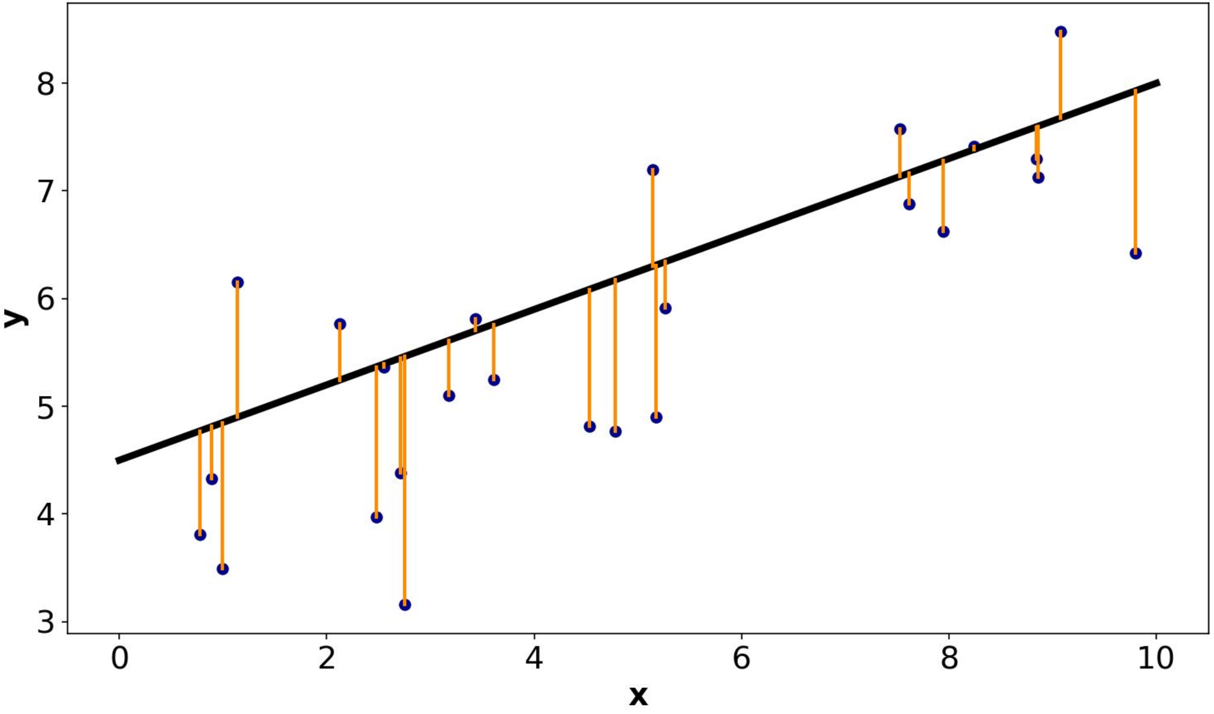
Linear Regression



Another possibility:

$y = 4.5 + 0.35x$

Linear Regression



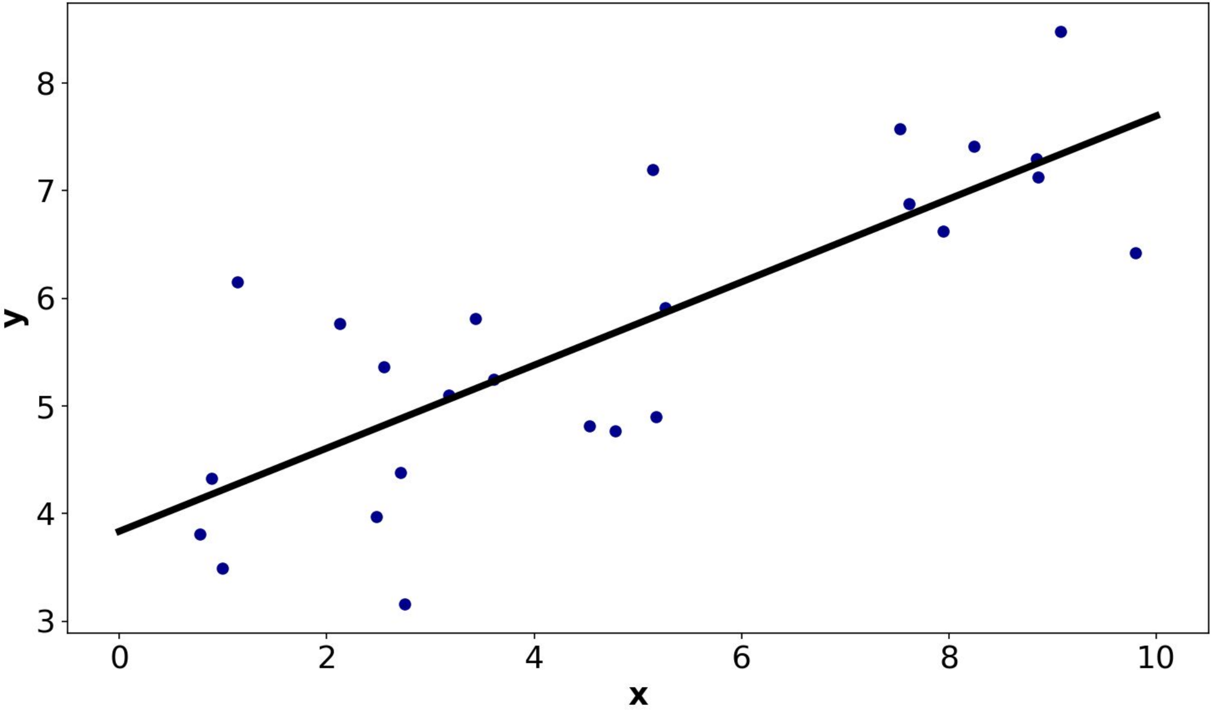
Another possibility:

$y = 4.5 + 0.35x$

Here,

RSS = 24.28

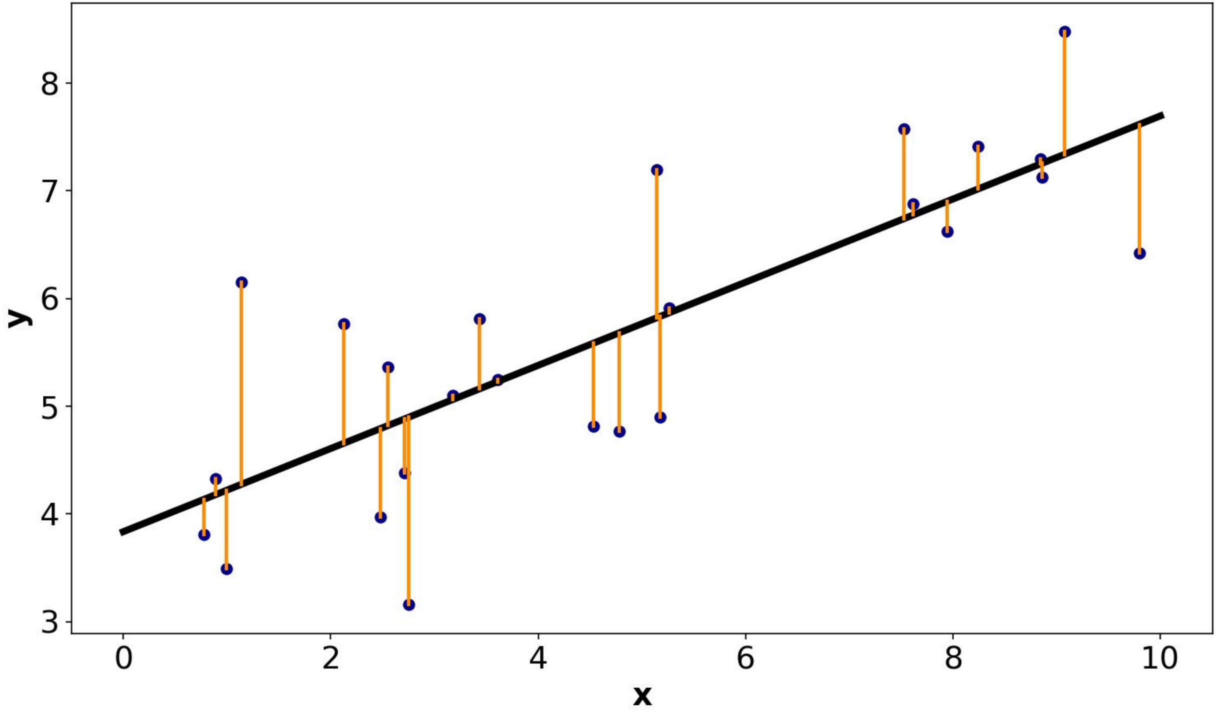
Linear Regression



The best possible:

$y = 3.84 + 0.386x$

Linear Regression



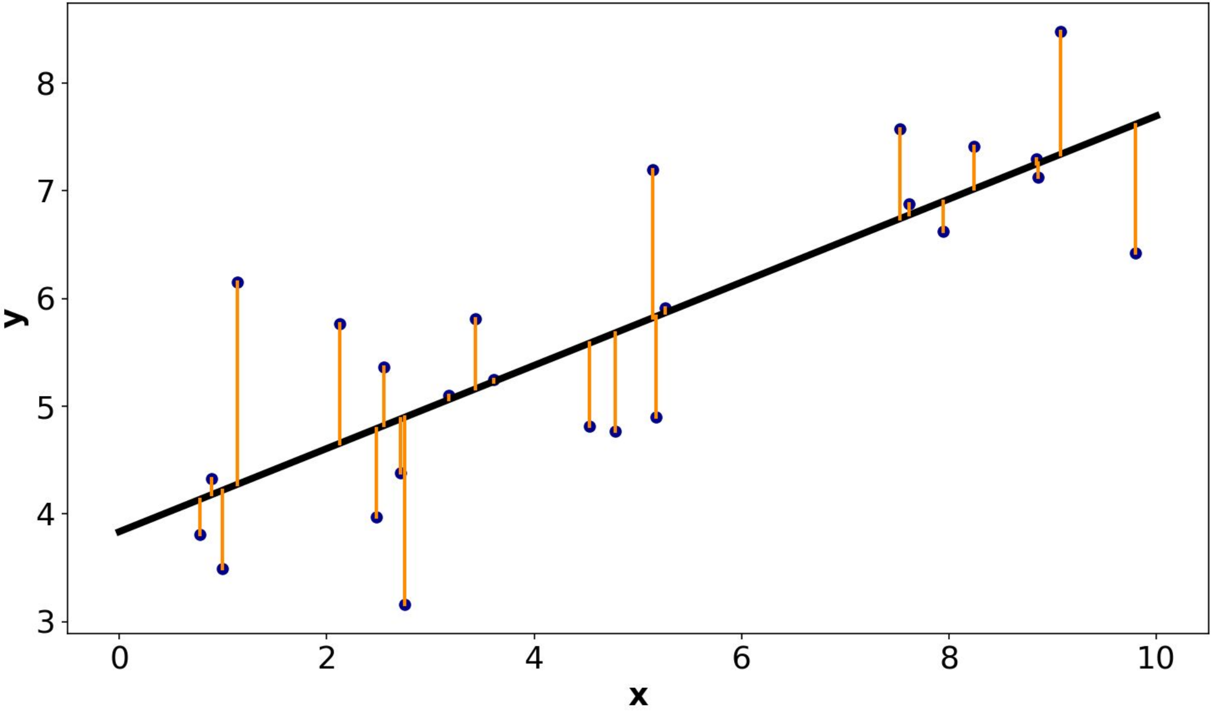
The best possible:

$y = 3.84 + 0.386x$

Here,

RSS = 17.97

Linear Regression



For the best-fitting line, the average (absolute) residual is equal to 0.67.

Can we expect that on new data generated by the same process, we will be off on average by 0.67 still?