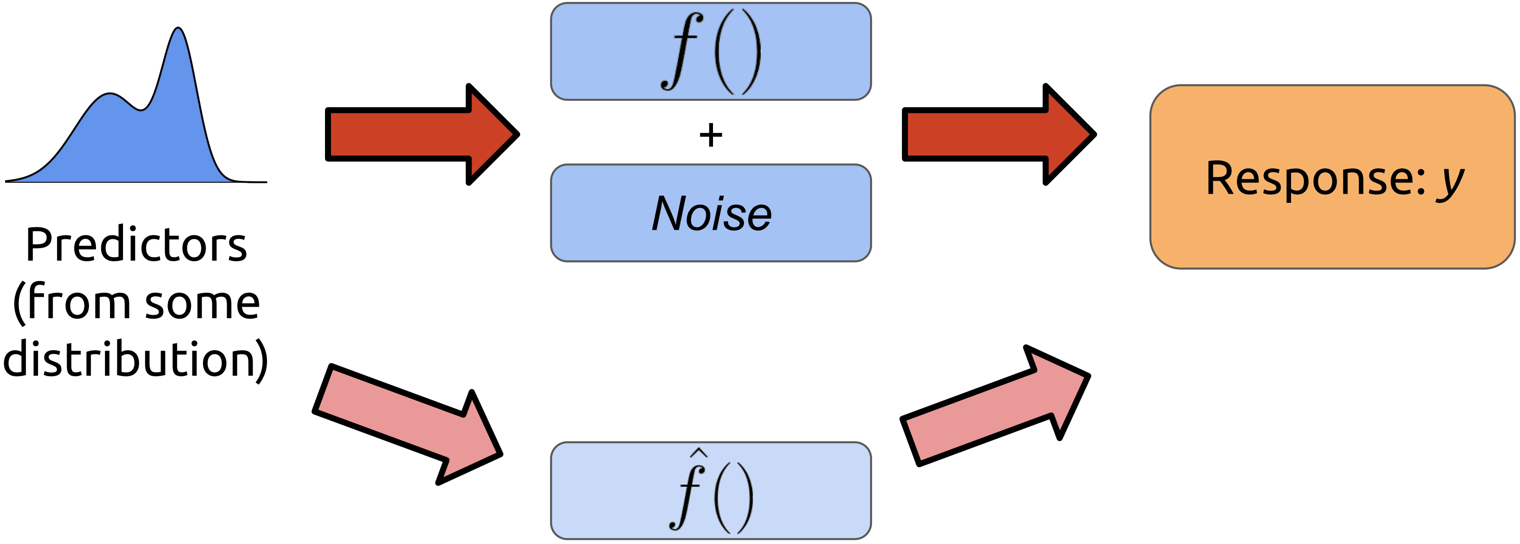
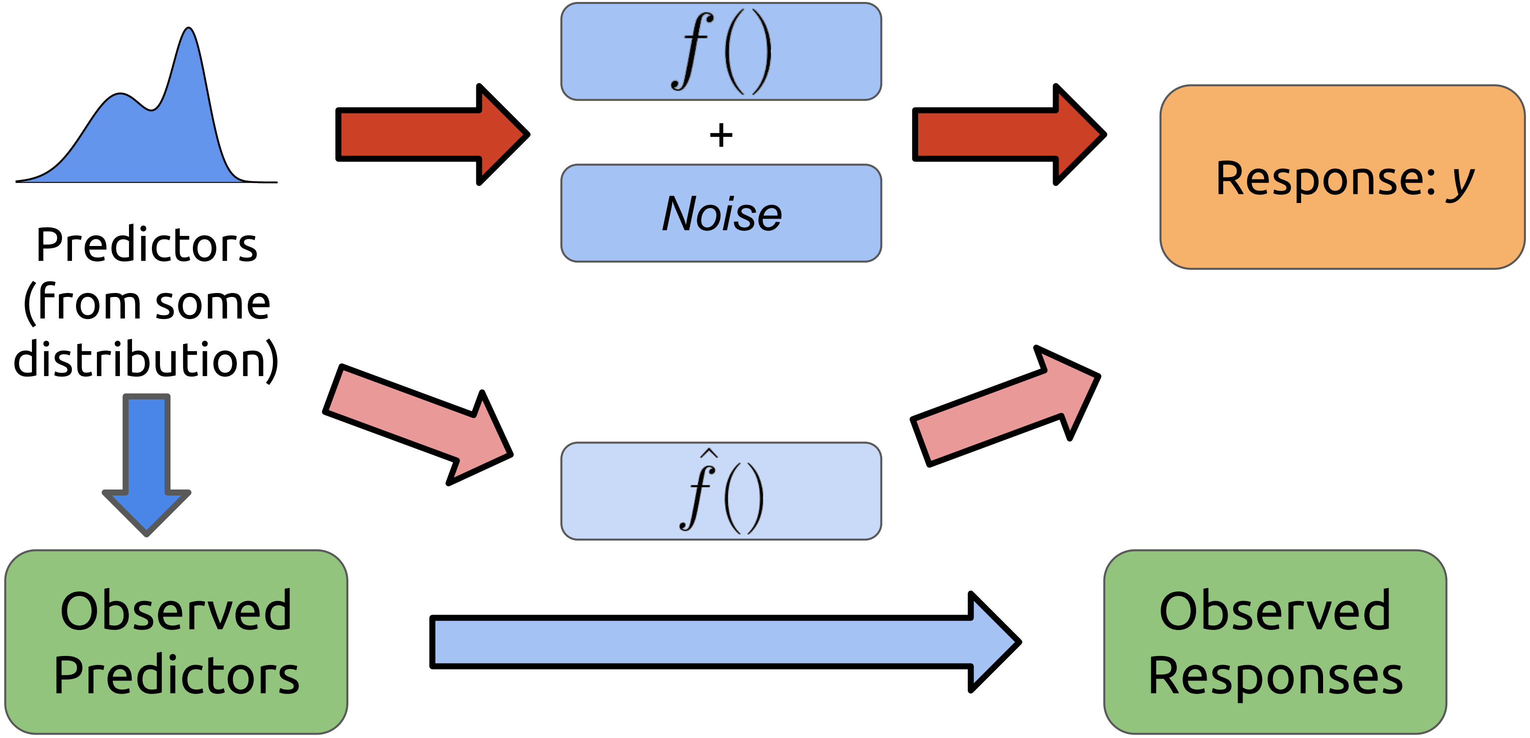
Introduction to Linear Regression

Supervised Learning - How



Supervised Learning - How



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 $\hat{f}()$

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 use a particularly simple functional form to make predictions - a weighted sum of the predictor variables.

Linear Regression - Example

Let’s say we want to build a model to predict home price.

Starting simple, let’s say we know the square footage of living space and the price of a set of homes.

Linear Regression - Example

	id	sqft_living	price
0	7129300520	1180	221900.0
1	6414100192	2570	538000.0
2	5631500400	770	180000.0
3	2487200875	1960	604000.0
4	1954400510	1680	510000.0
5	7237550310	5420	1225000.0
6	1321400060	1715	257500.0
7	2008000270	1060	291850.0
8	2414600126	1780	229500.0
9	3793500160	1890	323000.0

Here's a sample from our observed data.

Linear Regression - Example

Predictors	**Target**
sqft_living	price

Linear Regression - Example

Predictors	**Target**
sqft_living	price

Approach 1: Multiply sqft_living by a constant to predict price.

$$\text{predicted price} = \boxed{} \cdot \text{sqft\_living}$$

Linear Regression - Example

Predictors	**Target**
sqft_living	price

Approach 1: Multiply sqft_living by a constant to predict price.

$$\text{predicted price} = \boxed{} \cdot \text{sqft\_living}$$

Determine what goes here
based on the observed data.

Linear Regression - Example

Predictors	**Target**
sqft_living	price

Approach 1.5: Start with a “base price” and then add sqft_living multiplied by a constant to predict price.

$$\text{predicted price} = \boxed{} + \boxed{} \cdot \text{sqft\_living}$$

Determine what goes here
based on the observed data.

Linear Regression

How do we find the values for this coefficient?

Linear Regression

How do we find the values for this coefficient?

The approach we'll take for this example is to minimize the total squared **residuals** between the predicted and actual values for the data used to fit/train the model.

Linear Regression

How do we find the values for this coefficient?

The approach we'll take for this example is to minimize the total squared **residuals** between the predicted and actual values for the data used to fit/train the model.

True values

residual for observation *i*: $y_i - \hat{f}(\vec{x}_i)$

Predicted Values

Linear Regression

How do we find the values for this coefficient?

The approach we'll take for this example is to minimize the total squared **residuals** between the predicted and actual values for the data used to fit/train the model.

True values

squared residual for observation *i*: $$(y_i - \hat{f}(\vec{x}_i))^2$$

Predicted Values

Linear Regression

How do we find the values for this coefficient?

The approach we'll take for this example is to minimize the total squared **residuals** between the predicted and actual values for the data used to fit/train the model.

True values

total squared residuals:

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

Predicted Values

Linear Regression

How do we find the values for this coefficient?

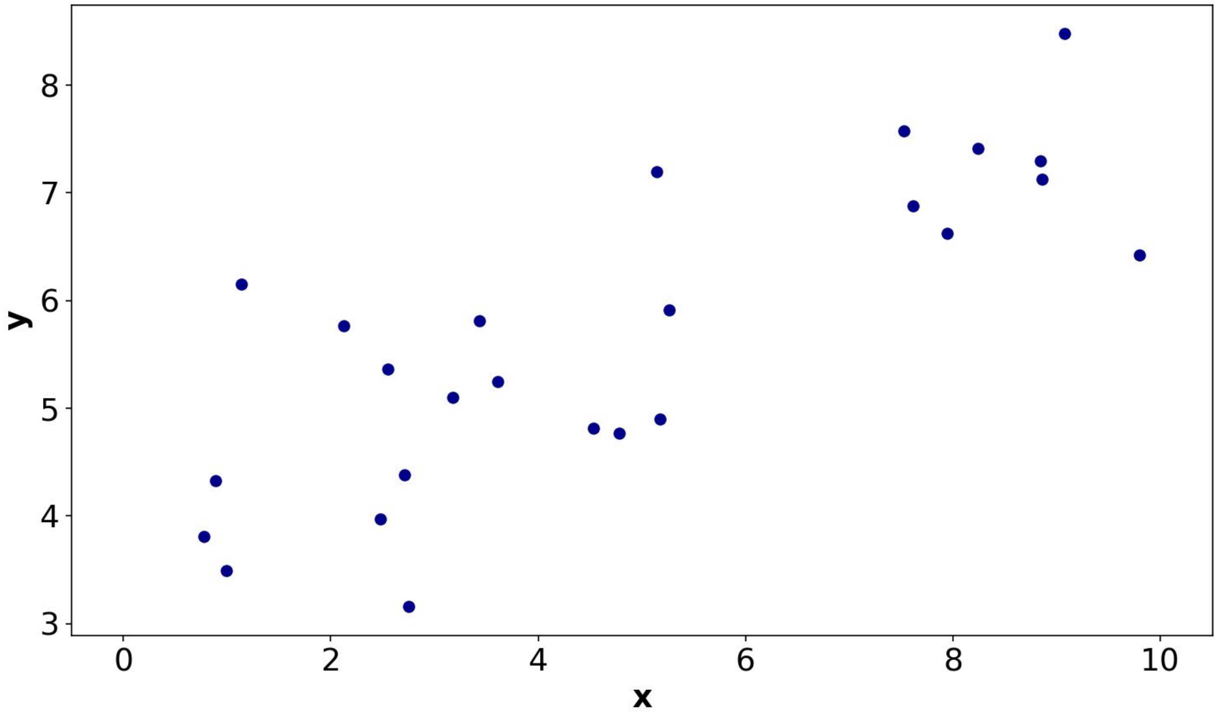
The approach we'll take for this example is to minimize the total squared **residuals** between the predicted and actual values for the data used to fit/train the model.

True values

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

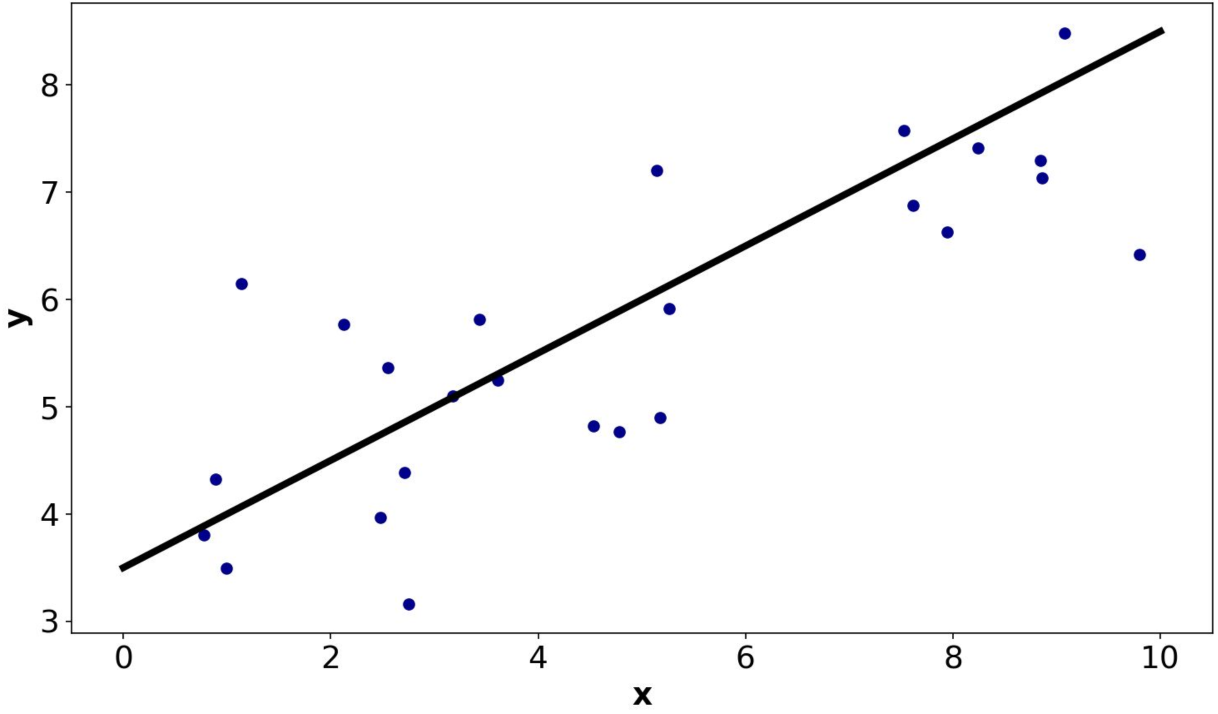
Predicted Values

Linear Regression



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

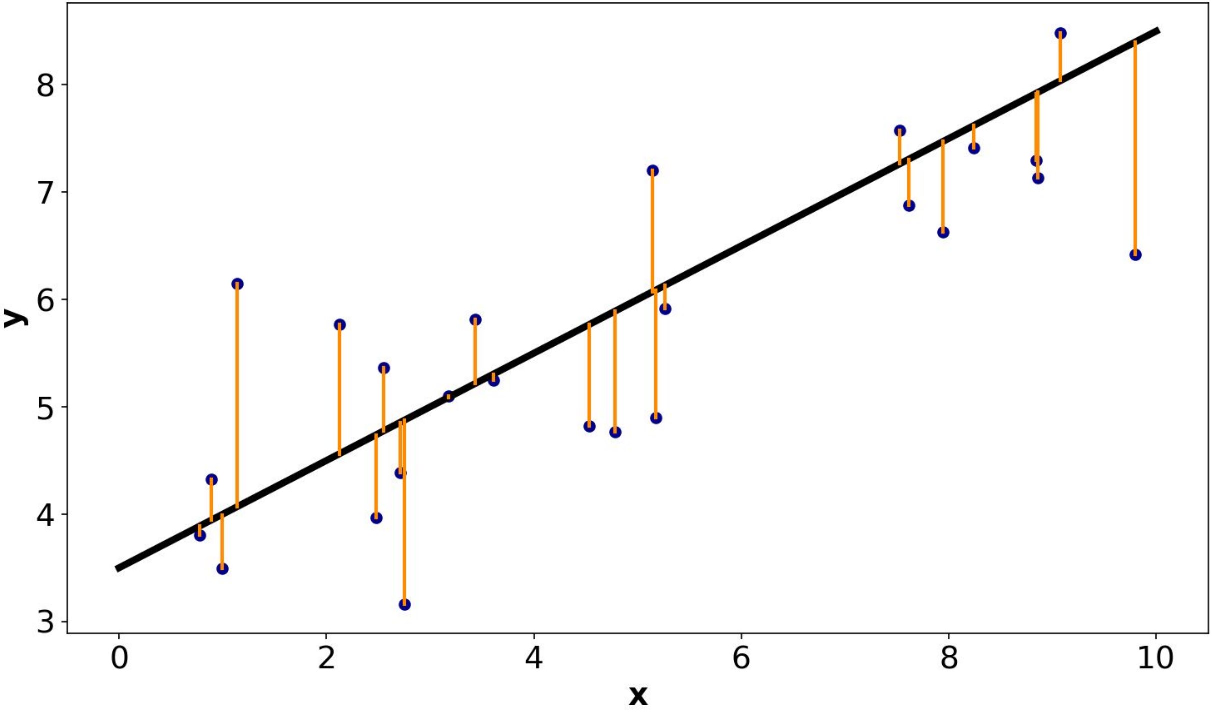
Linear Regression



One possible line:

y = 3.5 + 0.5x

Linear Regression



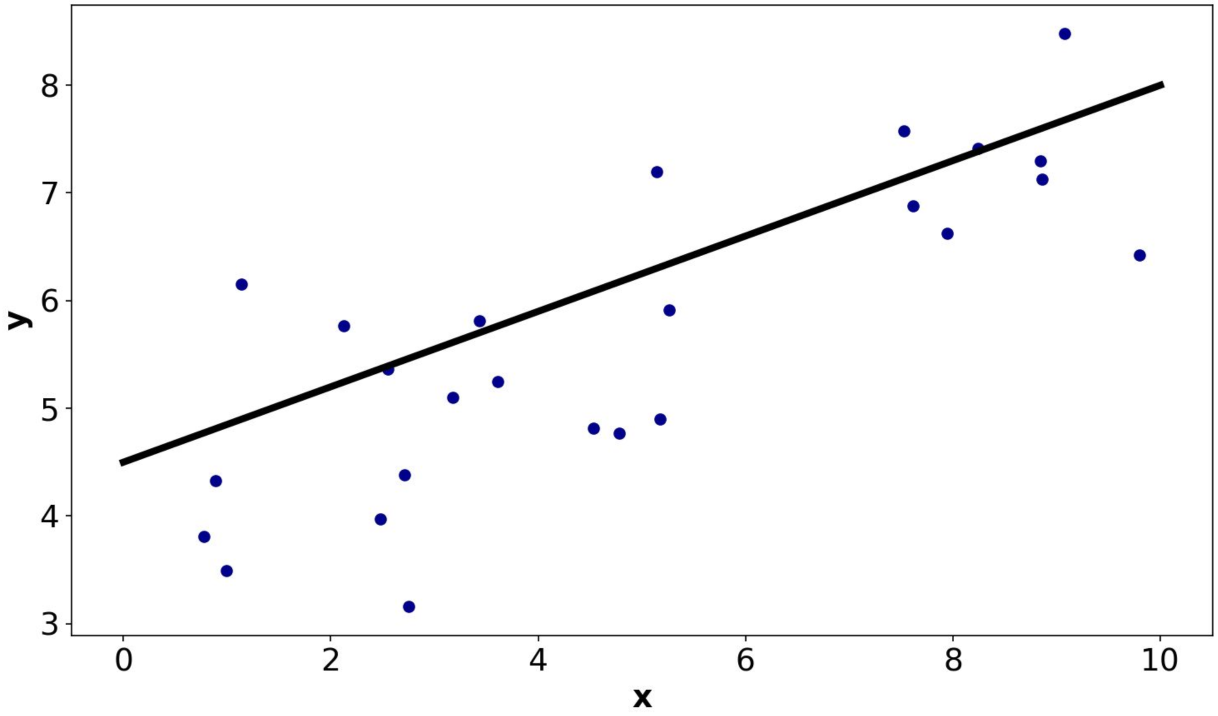
One possible line:

y = 3.5 + 0.5x

For this line,

RSS = 20.36

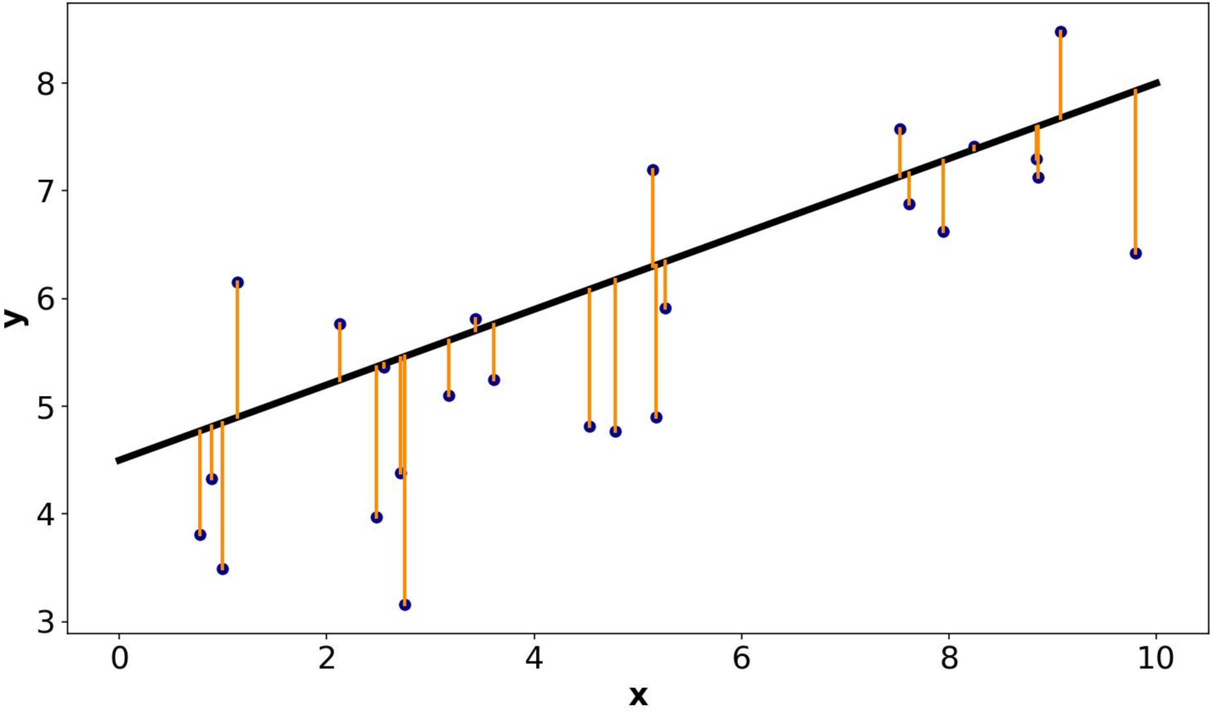
Linear Regression



Another possibility:

$y = 4.5 + 0.35x$

Linear Regression



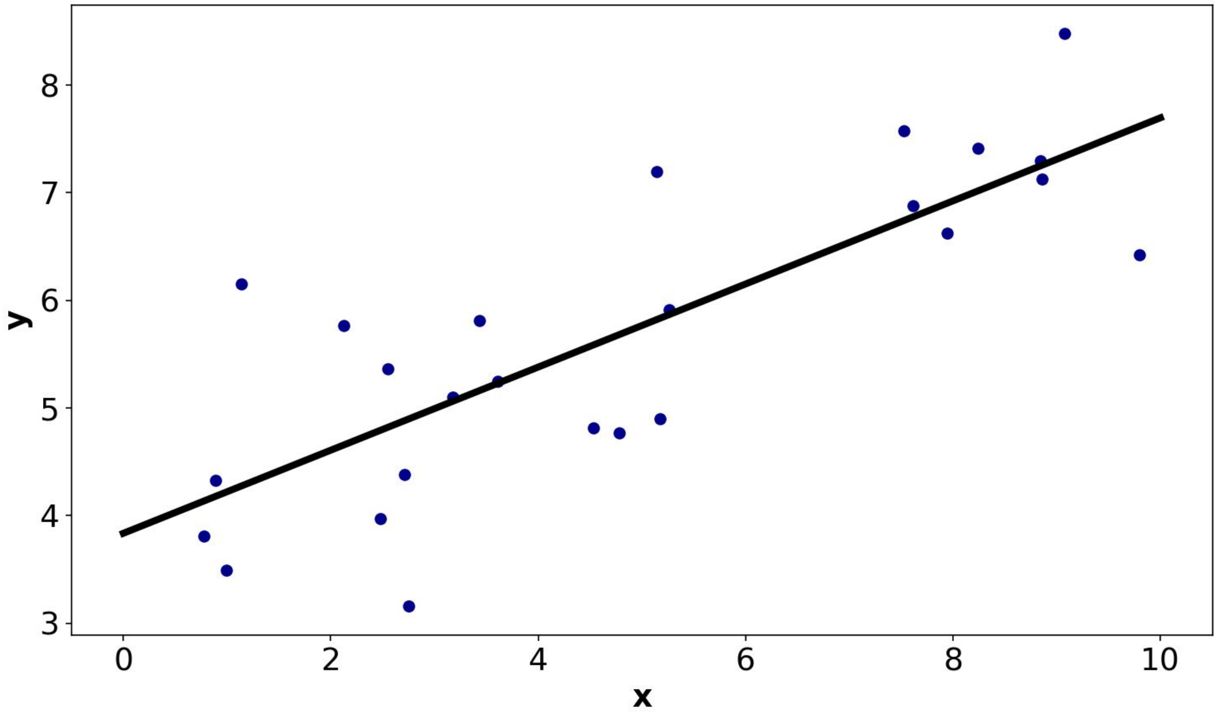
Another possibility:

y = 4.5 + 0.35x

Here,

RSS = 24.28

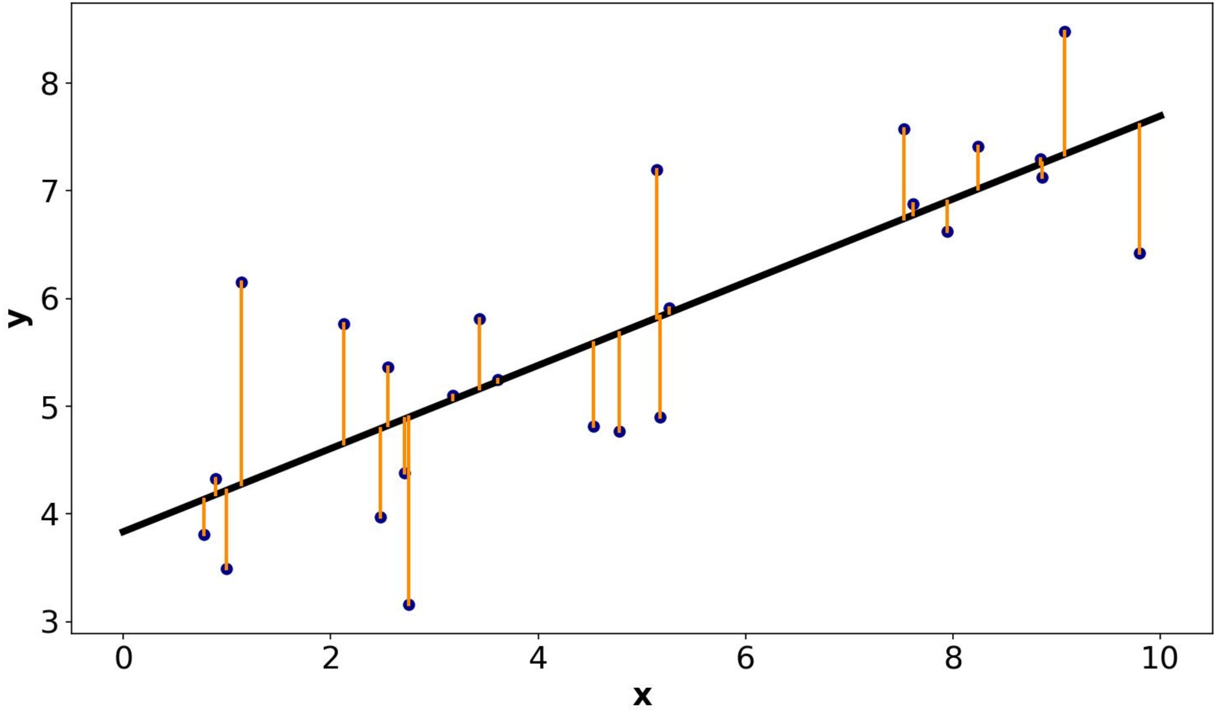
Linear Regression



The best possible:

$y = 3.84 + 0.386x$

Linear Regression



The best possible:

$y = 3.84 + 0.386x$

Here,

RSS = 17.97

Linear Regression

How *exactly* do we minimize RSS?

Linear Regression

How *exactly* do we minimize RSS?

Through some kind of optimization algorithm:

- Analytical solution could be used in this case (using matrix algebra tricks)
- Limited-memory BFGS (https://en.wikipedia.org/wiki/Limited-memory_BFGS)
- Gradient descent (https://en.wikipedia.org/wiki/Gradient_descent)

Linear Regression

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

If we have a lot of observed points, we might instead use **mean squared error (MSE)**.

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2}{n}$$

Minimizing RSS is equivalent to minimizing MSE (Why?)

Supervised Learning - Goals

Very Important Note: For linear regression, we find the coefficients by minimizing RSS/MSE on the training data, but this does not guarantee that the model will generalize well.

It is important to do a train/test split and estimate the *generalization error* - the only thing that we care about in evaluating a machine learning model.

Linear Regression - Example

Predictors	**Target**
sqft_living	price

Approach 1.5: Start with a “base price” and then add sqft_living multiplied by a constant to predict price.

$$\text{predicted price} = \boxed{} + \boxed{} \cdot \text{sqft\_living}$$

Determine what goes here based on the observed data.

Linear Regression - Example

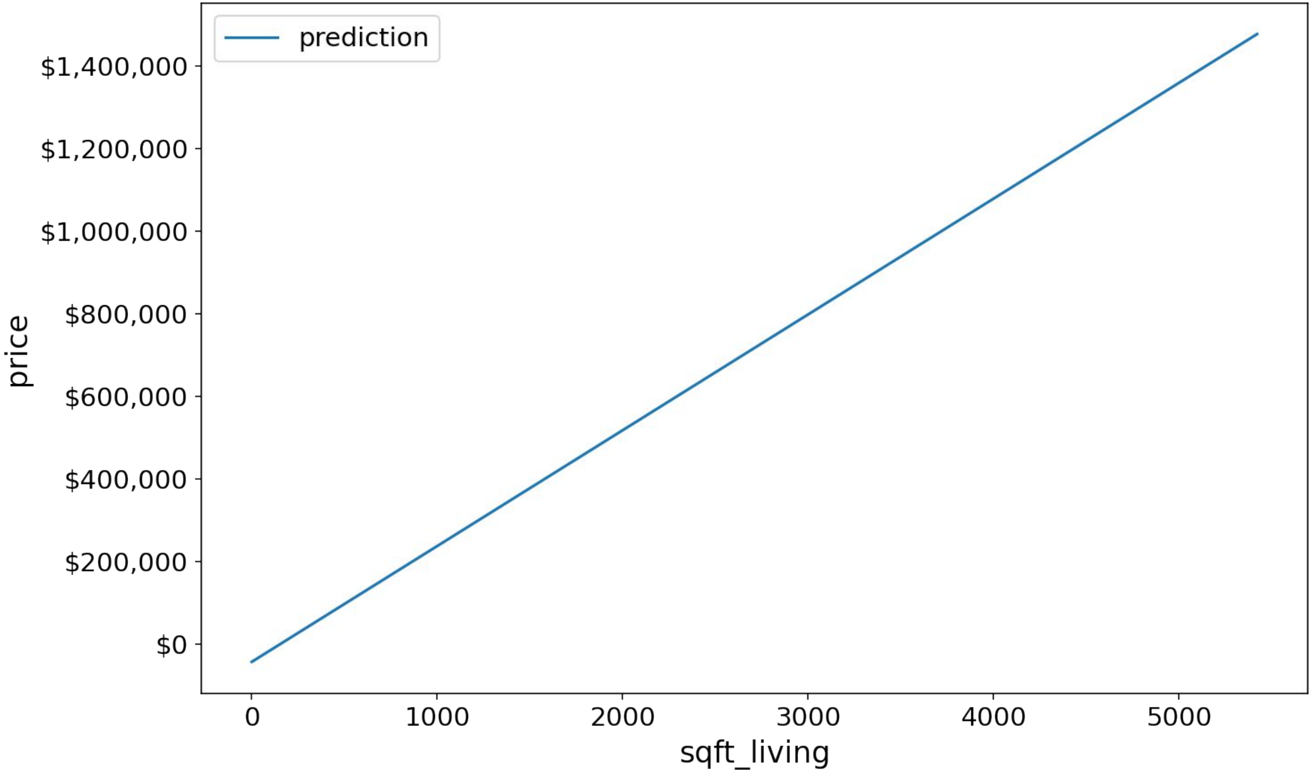
Predictors	**Target**
sqft_living	price

Approach 1.5: Start with a “base price” and then add sqft_living multiplied by a constant to predict price.

$$\text{predicted price} = \boxed{-42123} + \boxed{281} \cdot \text{sqft\_living}$$

Determine what goes here based on the observed data.

Linear Regression - Example



Linear Regression - Example

	id	sqft_living	price	predicted_price
0	7129300520	1180	221900.0	289118.0
1	6414100192	2570	538000.0	679309.0
2	5631500400	770	180000.0	174026.0
3	2487200875	1960	604000.0	508074.0
4	1954400510	1680	510000.0	429474.0
5	7237550310	5420	1225000.0	1479340.0
6	1321400060	1715	257500.0	439299.0
7	2008000270	1060	291850.0	255433.0
8	2414600126	1780	229500.0	457546.0
9	3793500160	1890	323000.0	488424.0

Applying this to our data gives these results.

Linear Regression - Example

	id	sqft_living	price	predicted_price
0	7129300520	1180	221900.0	289118.0

Let’s see how we arrived at this predicted price.

Linear Regression - Example

	id	sqft_living	price	predicted_price
0	7129300520	1180	221900.0	289118.0

Let’s see how we arrived at this predicted price.

$$\text{predicted price} = -42123 + 281 \cdot \text{sqft\_living}$$

Linear Regression - Example

	id	sqft_living	price	predicted_price
0	7129300520	1180	221900.0	289118.0

Let’s see how we arrived at this predicted price.

$$\text{predicted price} = -42123 + 281 \cdot \text{sqft\_living}$$

$$\text{predicted price} = -42123 + 281 \cdot (1180)$$

Linear Regression - Example

	id	sqft_living	price	predicted_price
0	7129300520	1180	221900.0	289118.0

Let’s see how we arrived at this predicted price.

$$\text{predicted price} = -42123 + 281 \cdot \text{sqft\_living}$$

$$\text{predicted price} = -42123 + 281 \cdot (1180)$$

$$\text{predicted price} = -42123 + 331580$$

Linear Regression - Example

	id	sqft_living	price	predicted_price
0	7129300520	1180	221900.0	289118.0

Let's see how we arrived at this predicted price.

$$\text{predicted price} = -42123 + 281 \cdot \text{sqft\_living}$$

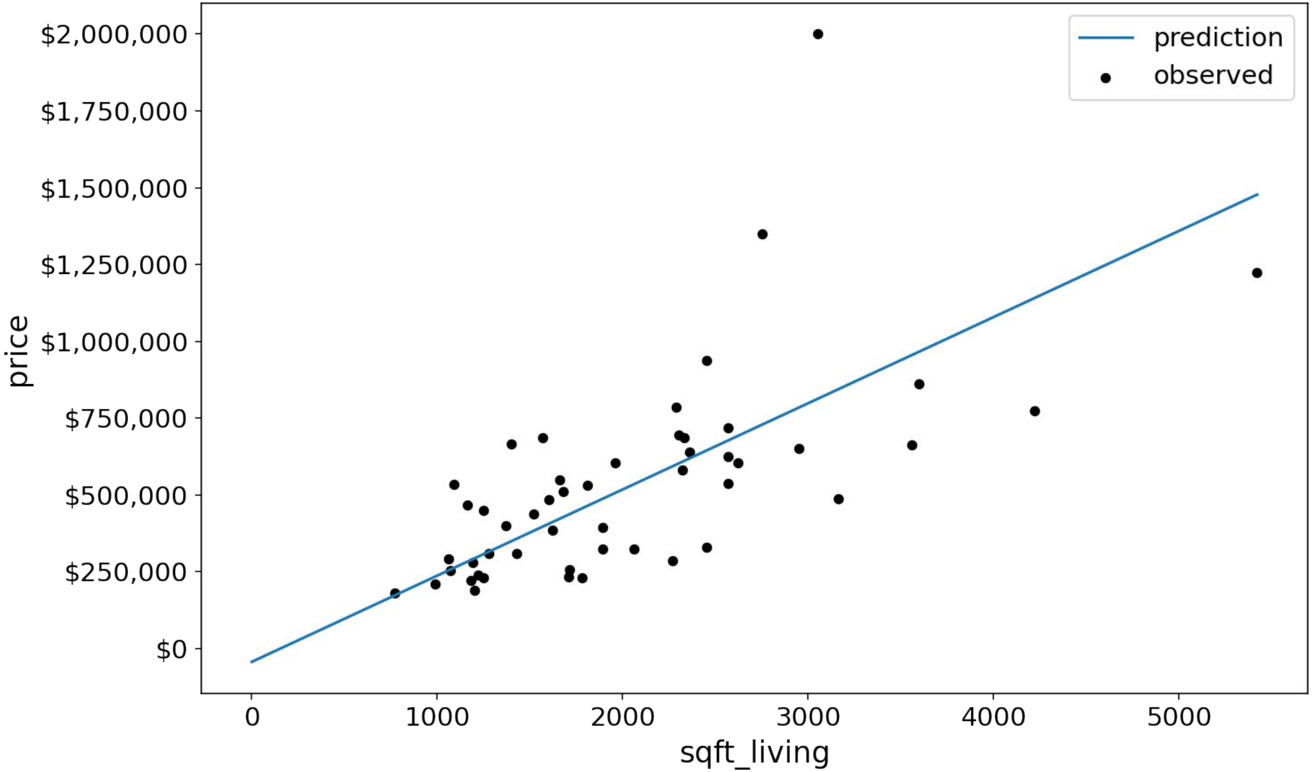
$$\text{predicted price} = -42123 + 281 \cdot (1180)$$

$$\text{predicted price} = -42123 + 331580$$

$$\text{predicted price} = 289457^{*}$$

* This number is slightly different due to rounding the coefficients

Linear Regression - Example



Linear Regression - Example

This looks okay, but could potentially be improved. What if we add in more information about our observations?

Linear Regression - Example

	id	sqft_living	condition	price
0	7129300520	1180	3	221900.0
1	6414100192	2570	3	538000.0
2	5631500400	770	3	180000.0
3	2487200875	1960	5	604000.0
4	1954400510	1680	3	510000.0
5	7237550310	5420	3	1225000.0
6	1321400060	1715	3	257500.0
7	2008000270	1060	3	291850.0
8	2414600126	1780	3	229500.0
9	3793500160	1890	3	323000.0

This looks okay, but could potentially be improved. What if we add in more information about our observations?

Linear Regression - Example

<u>Predictors</u>

sqft_living
condition

<u>Target</u>

price

$$\text{predicted price} = \boxed{} + \boxed{} \cdot \text{sqft\_living} + \boxed{} \cdot \text{condition}$$

Determine what goes here
based on the observed data.

Linear Regression - Example

Predictors

sqft_living
condition

Target

price

$$\text{predicted price} = \boxed{-192628} + \boxed{282} \cdot \text{sqft\_living} + \boxed{43067} \cdot \text{condition}$$

Determine what goes here
based on the observed data.

Linear Regression - Example

$$\text{predicted price} = -192628 + 282 \cdot \text{sqft\_living} + \boxed{43067} \cdot \text{condition}$$

How do we interpret
this value?

Linear Regression - Example

$$\text{predicted price} = -192628 + 282 \cdot \text{sqft\_living} + \boxed{43067} \cdot \text{condition}$$

How do we interpret
this value?

If two houses have the same sqft_living, for every one unit difference in condition, their predicted prices will be differ by $43,067.

Linear Regression - Example

$$\text{predicted price} = -192628 + \boxed{282} \cdot \text{sqft\_living} + 43067 \cdot \text{condition}$$

How do we interpret
this value?

Linear Regression - Example

$$\text{predicted price} = -192628 + \boxed{282} \cdot \text{sqft\_living} + 43067 \cdot \text{condition}$$

How do we interpret
this value?

If two houses have the same condition for every unit of difference in sqft_living, their predicted prices will differ by $282.

Linear Regression - Example

What if we have even more predictors?

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.