Introduction to Linear Regression

SML 312 – Fall 2022

What is Regression?

What is Regression?

- **Regression** predicts/explains one continuous variable in terms of other continuous variables.

What is Linear Regression?

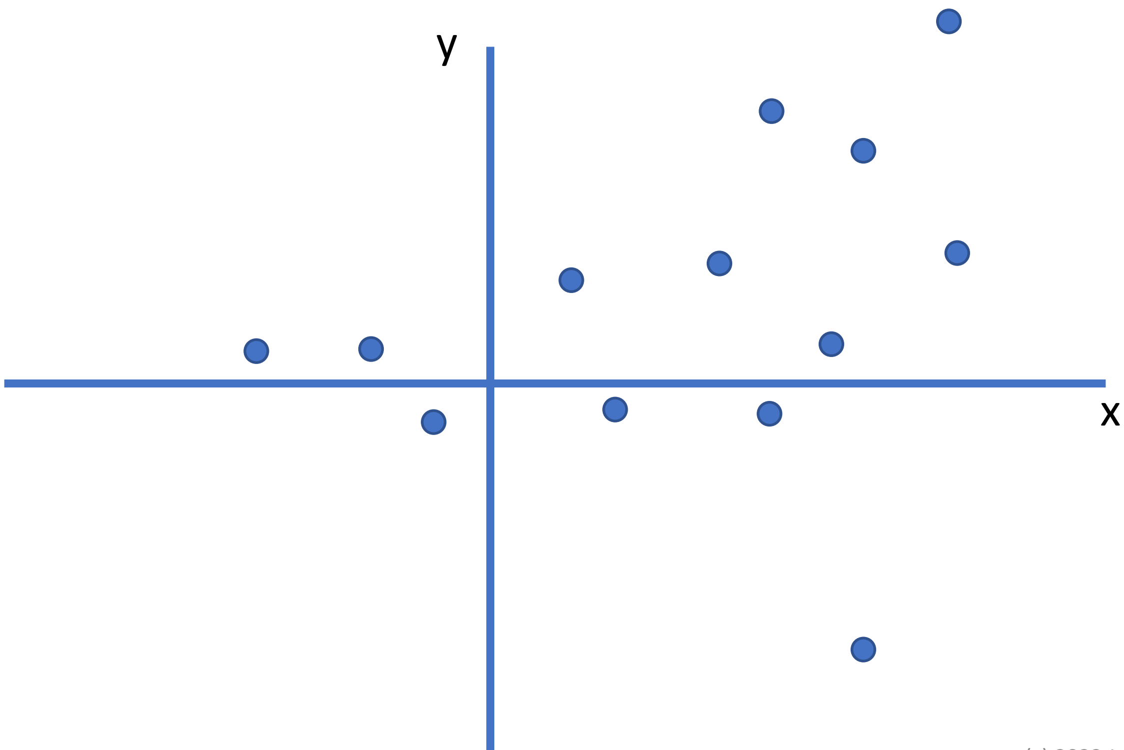
- **Regression** predicts/explains one continuous variable in terms of other continuous variables.

What is Linear Regression?

- **Regression** predicts/explains one continuous variable in terms of other continuous variables.
- **Linear Regression** does this by using a linear model to perform this prediction.

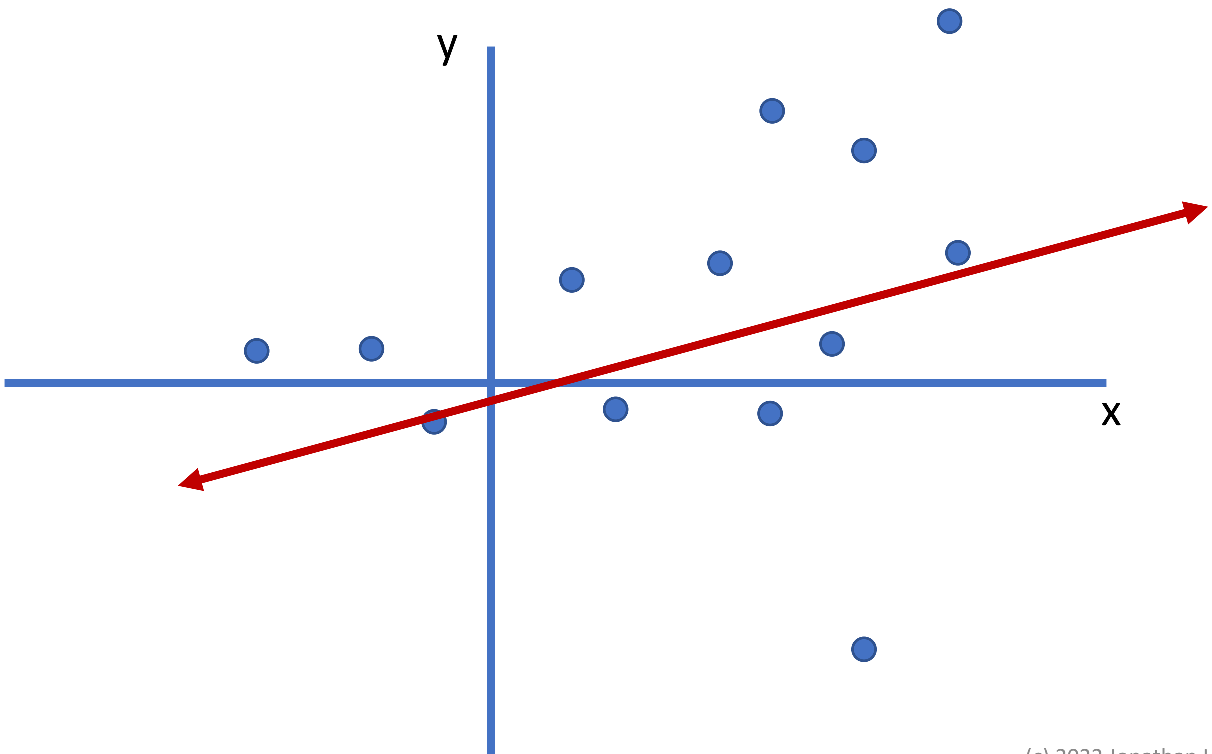
What is Linear Regression?

- **Regression** predicts/explains one continuous variable in terms of other continuous variables.
- **Linear Regression** does this by using a linear model to perform this prediction.



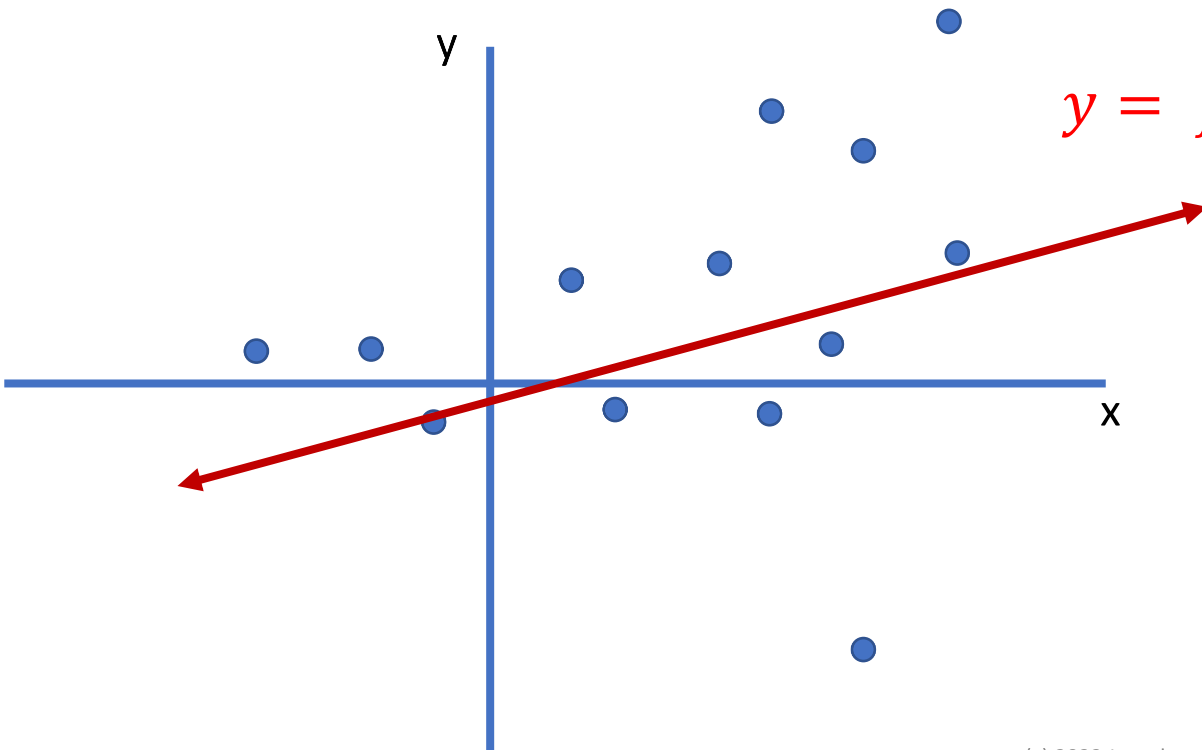
What is Linear Regression?

- **Regression** predicts/explains one continuous variable in terms of other continuous variables.
- **Linear Regression** does this by using a linear model to perform this prediction.



What is Linear Regression?

- **Regression** predicts/explains one continuous variable in terms of other continuous variables.
- **Linear Regression** does this by using a linear model to perform this prediction.



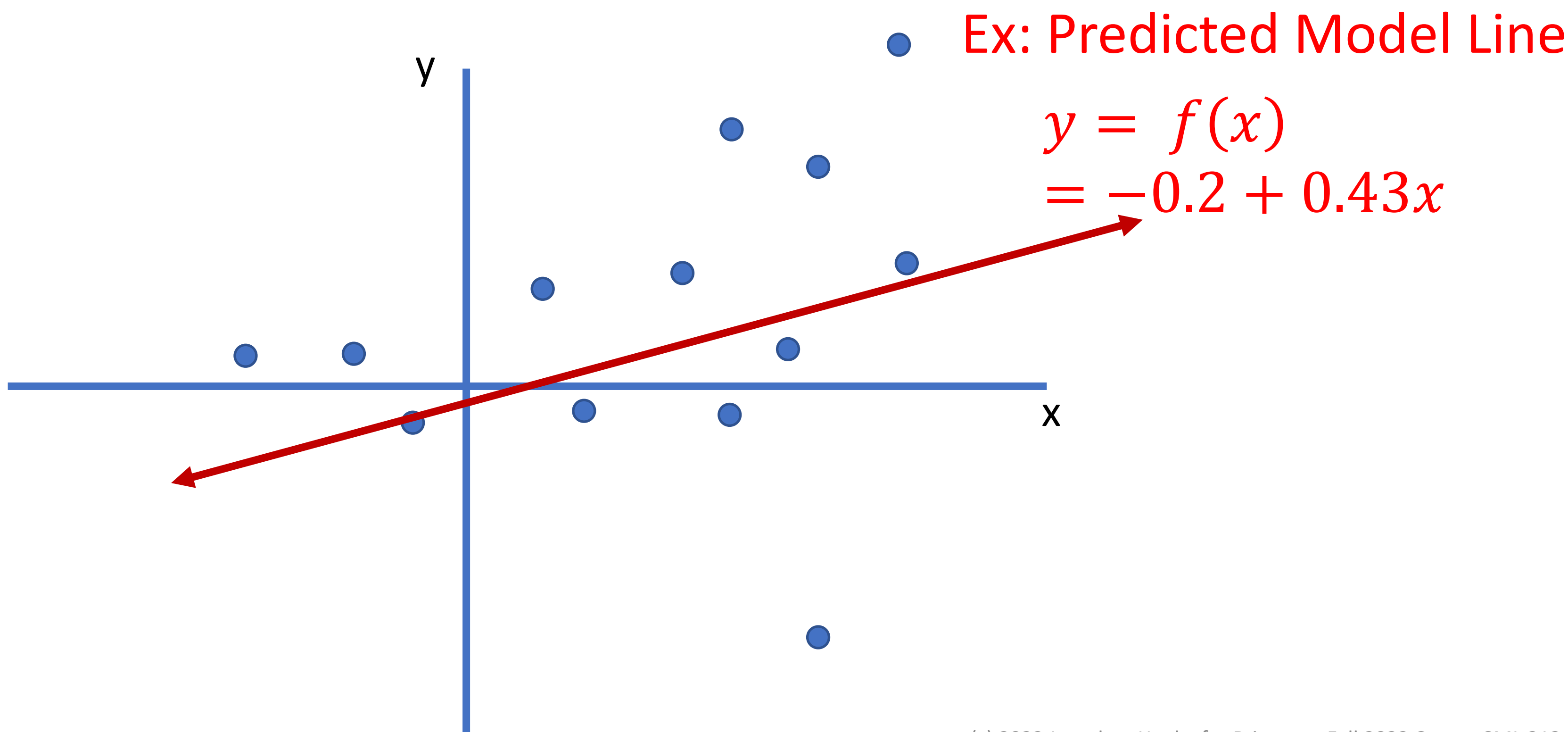
$$y = f(x) = b_0 + b_1 x$$

Ex:

$$b_0 = -0.2$$
$$b_1 = 0.43$$

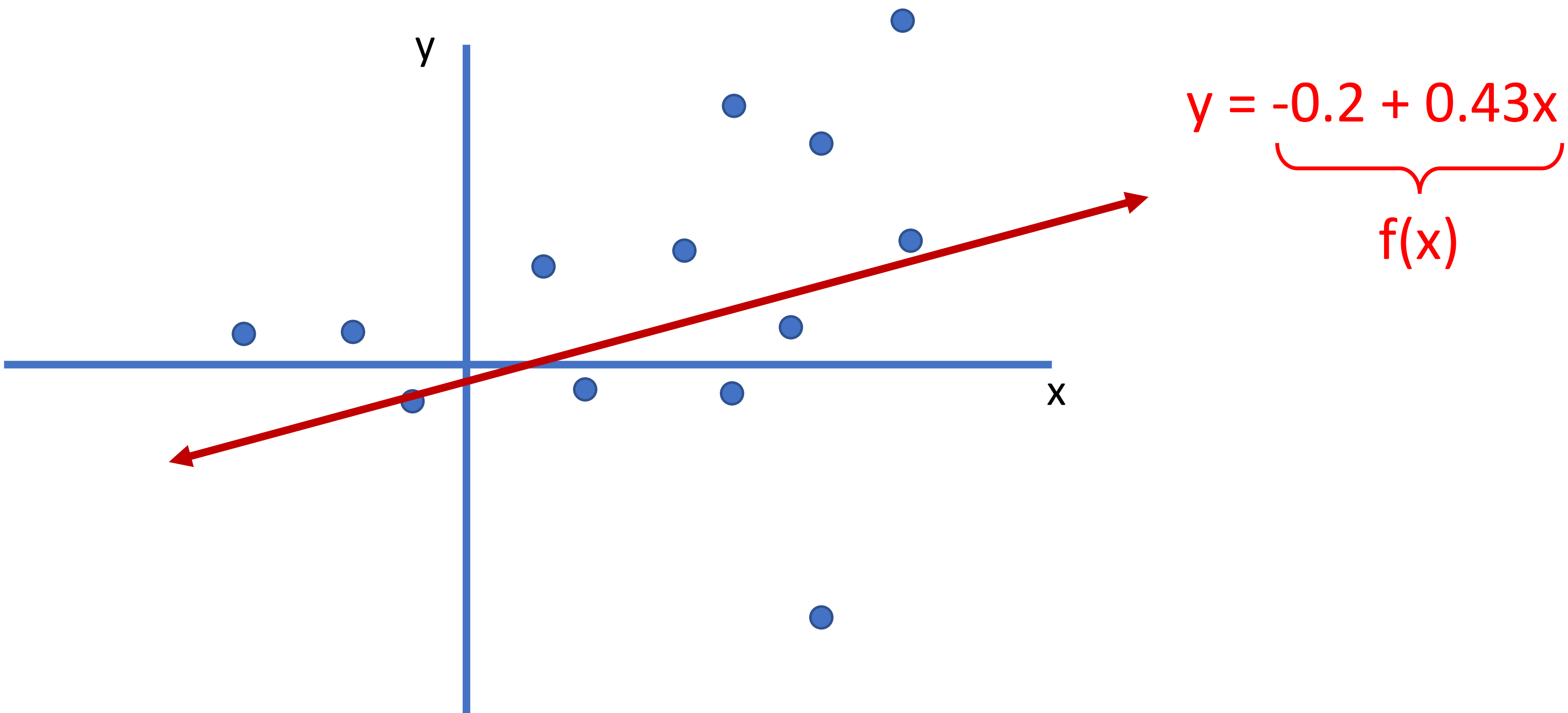
What is Linear Regression?

- **Regression** predicts/explains one continuous variable in terms of other continuous variables.
- **Linear Regression** does this by using a linear model to perform this prediction.

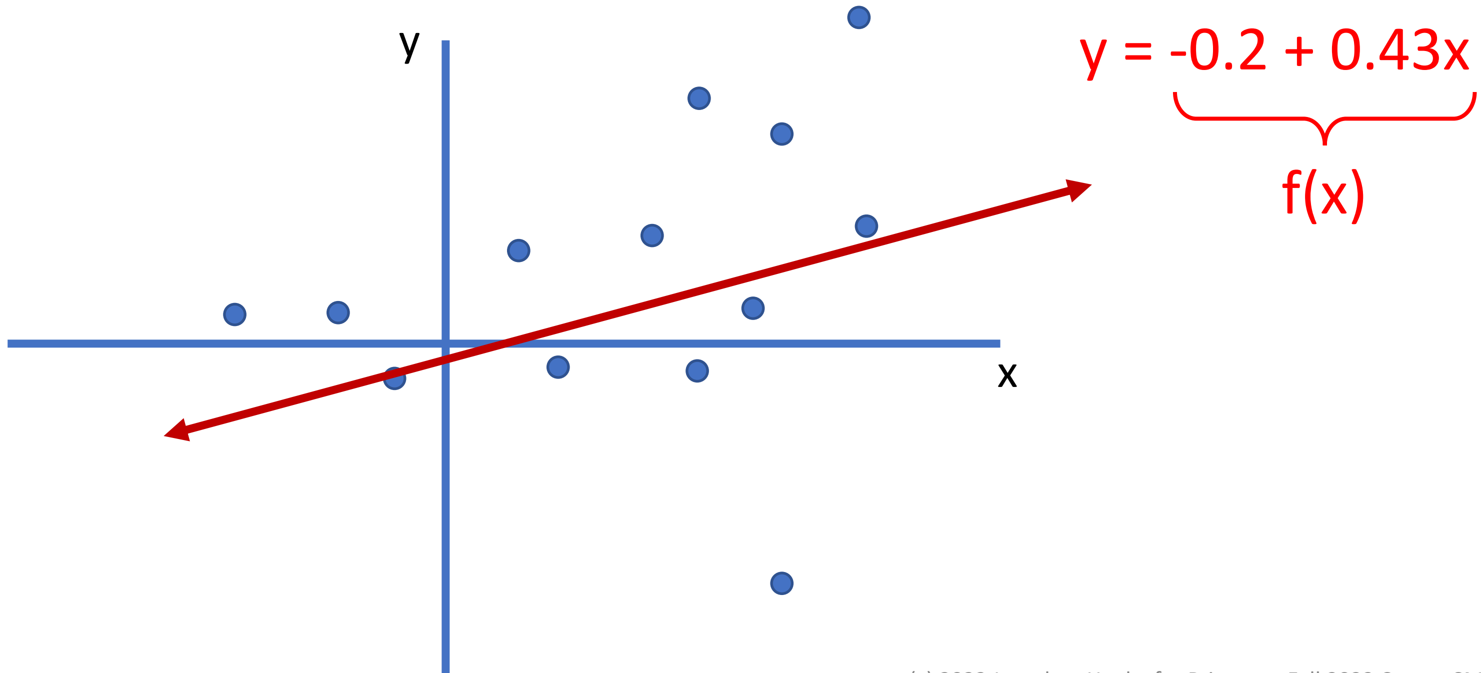


What is Linear Regression?

- In particular, when we have a <u>single feature</u> (i.e. x), Linear regression answers the question: What is the "best" line through a set of data points $\{(x^{(i)}, y^{(i)})\}_{1 \le i \le n}$?

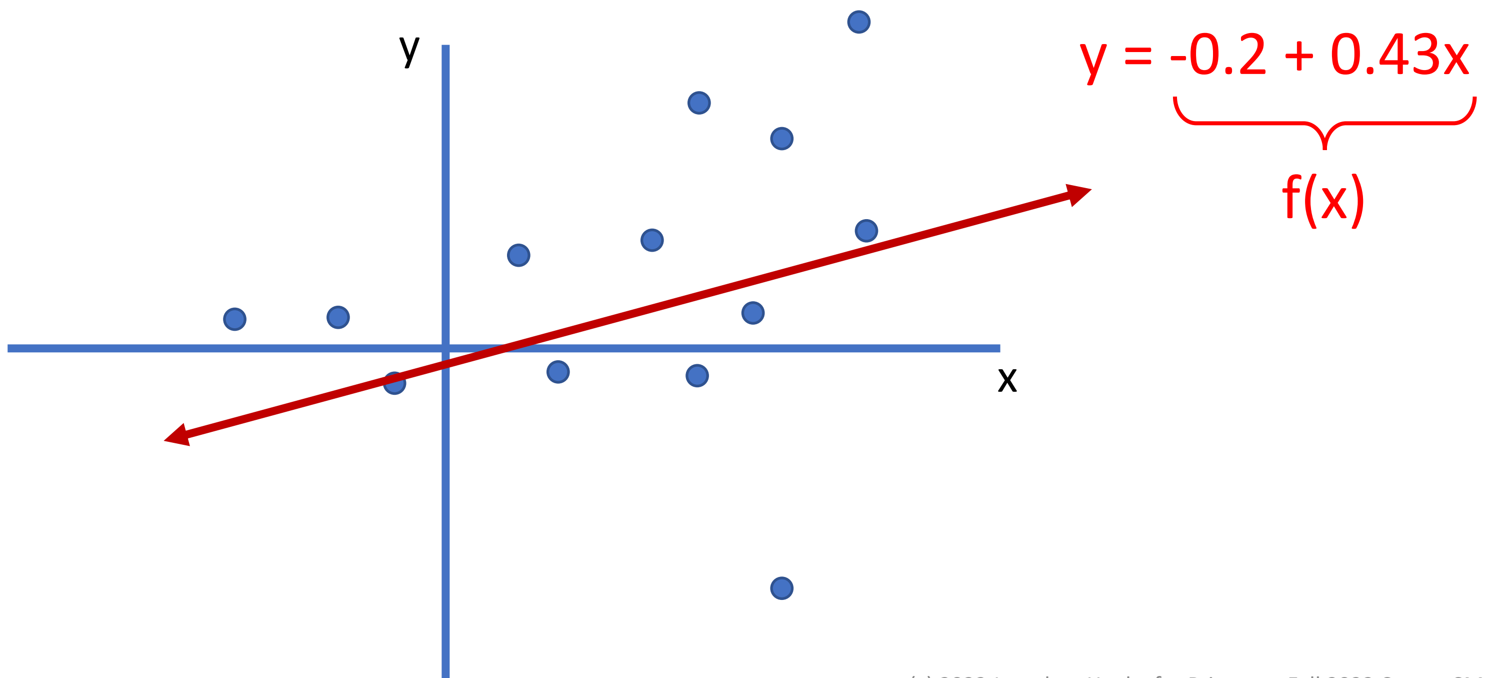


What do we mean by the "best" line?



What do we mean by the "best" line?

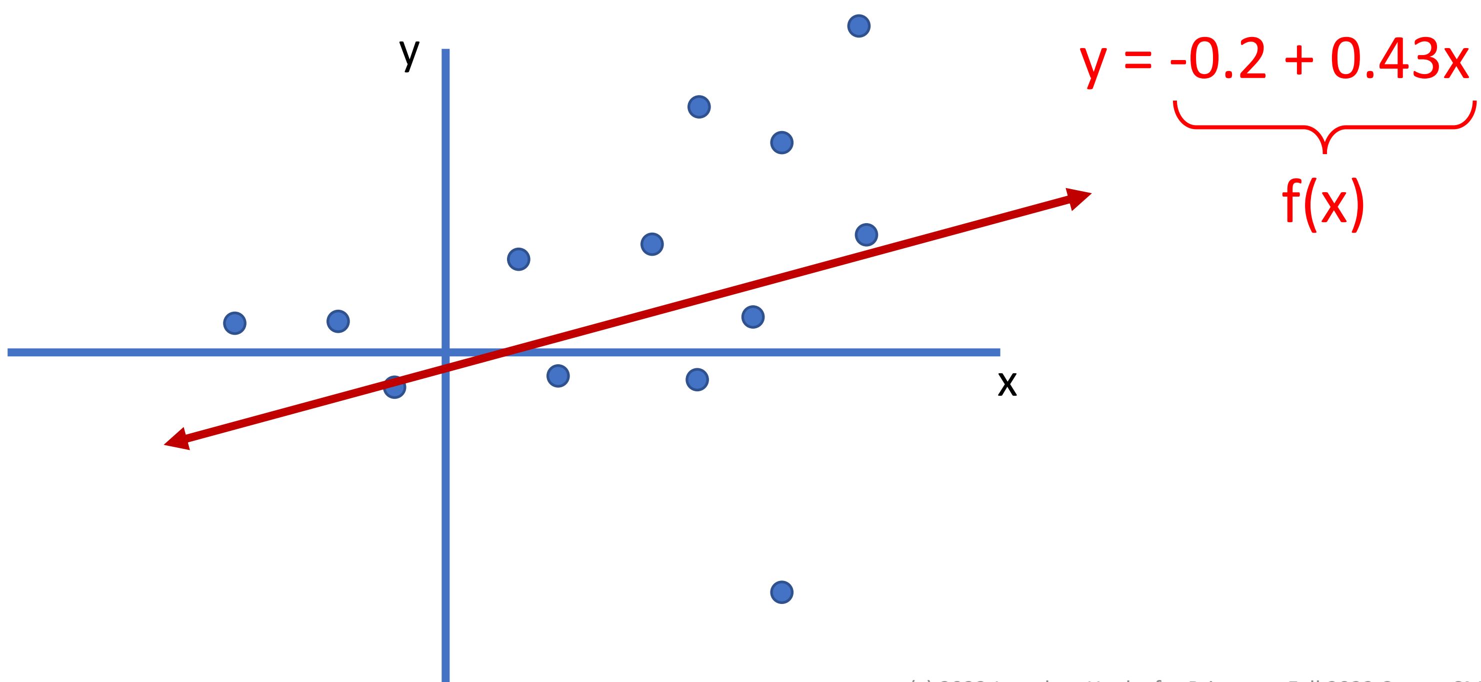
- The best line minimizes the (mean squared) error,



What do we mean by the "best" line?

- The best line minimizes the (mean squared) error, i.e. minimizes the number

Mean Squared Error (MSE) :=

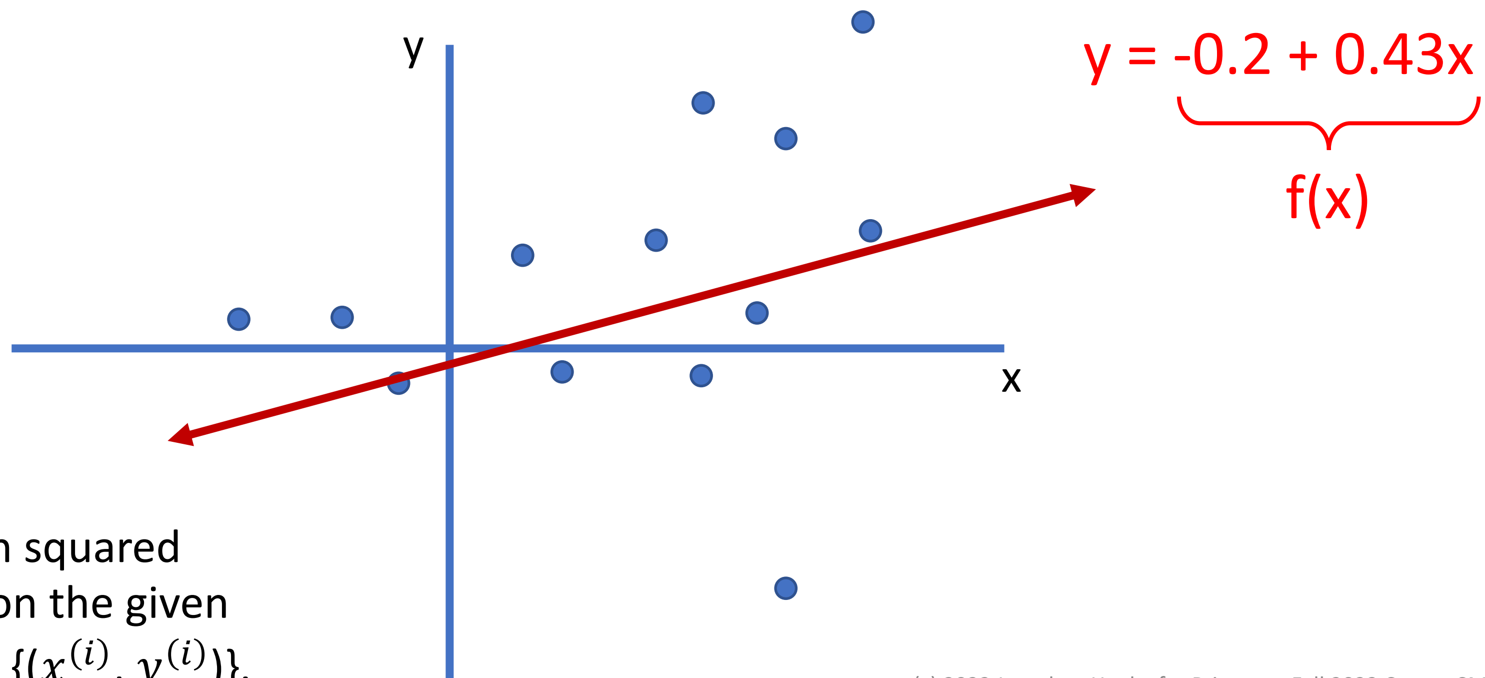


What do we mean by the “best” line?

- The best line minimizes the (mean squared) error, i.e. minimizes the number

Mean Squared Error (MSE) :=

$$\frac{1}{n}\left[\left(y^{(1)} - f\left(x^{(1)}\right)\right)^2 + \cdots + \left(y^{(n)} - f\left(x^{(n)}\right)\right)^2\right]$$



Note: The mean squared error depends on the given dataset of pairs $\{(x^{(i)}, y^{(i)})\}$.

What's the best (RV) linear predictor?

What's the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we're interested finding the "**best" linear predictor** of Y in terms of X.

What’s the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we’re interested finding the “**best” linear predictor** of Y in terms of X.

- This means that the expected Mean Square Error

$$\mathbb{E}[(Y - (b_0 + b_1 X))^2]$$

is minimized.

What's the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we're interested finding the "**best" linear predictor** of Y in terms of X.

- This means that the expected Mean Square Error

$$\mathbb{E}[(Y - (b_0 + b_1 X))^2]$$

is minimized.

- **Approach**: This can be done (using calculus!) to minimize the function MSE(b_0, b_1) as a function of the two parameters b_0 and b_1.

What's the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we're interested finding the "**best" linear predictor** of Y in terms of X.

- This means that the <u>expected</u> Mean Square Error

$$\mathbb{E}[(Y - (b_0 + b_1 X))^2]$$

  is minimized.

- **Approach:** This can be done (using calculus!) to minimize the function MSE(b_0, b_1) as a function of the two parameters b_0 and b_1.

- **Idea:** Expand this as a function (with numbers!) of b_0 and b_1. Then the minimum will occur when the derivatives (w.r.t. b_0 and b_1) are zero!

What’s the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we’re interested finding the “**best” linear predictor** of Y in terms of X.

- This means that the <u>expected</u> Mean Square Error

$$\mathbb{E}[(Y - (b_0 + b_1 X))^2]$$

  is minimized.

- **Approach:** This can be done (using calculus!) to minimize the function MSE(b_0, b_1) as a function of the two parameters b_0 and b_1.
- **Idea:** Expand this as a function (with numbers!) of b_0 and b_1. Then the minimum will occur when the derivatives (w.r.t. b_0 and b_1) are zero!

  - $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 x))^2]}{\partial b_0} =$

What’s the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we’re interested finding the “**best” linear predictor** of Y in terms of X.

- This means that the <u>expected</u> Mean Square Error

$$\mathbb{E}[(Y - (b_0 + b_1 X))^2]$$

is minimized.

- **Approach:** This can be done (using calculus!) to minimize the function MSE(b_0, b_1) as a function of the two parameters b_0 and b_1.
- **Idea:** Expand this as a function (with numbers!) of b_0 and b_1. Then the minimum will occur when the derivatives (w.r.t. b_0 and b_1) are zero!

    - $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 x))^2]}{\partial b_0} = \frac{\partial \mathbb{E}\left[Y^2 - 2(b_0 + b_1 x)Y + (b_0 + b_1 x)^2\right]}{\partial b_0}$

What's the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we're interested finding the "**best" linear predictor** of Y in terms of X.

- This means that the <u>expected</u> Mean Square Error

$$\mathbb{E}[(Y - (b_0 + b_1 X))^2]$$

  is minimized.

- **Approach:** This can be done (using calculus!) to minimize the function MSE(b_0, b_1) as a function of the two parameters b_0 and b_1.
- **Idea:** Expand this as a function (with numbers!) of b_0 and b_1. Then the minimum will occur when the derivatives (w.r.t. b_0 and b_1) are zero!

  - $$\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 x))^2]}{\partial b_0} = \frac{\partial \mathbb{E}[Y^2 - 2(b_0 + b_1 x)Y + (b_0 + b_1 x)^2]}{\partial b_0}$$
    $$= -2\mathbb{E}[Y] + 2b_0 + 2b_1 \mathbb{E}[X] = 0.$$

What’s the best (RV) linear predictor?

- Suppose we have two random variables, X and Y, and we’re interested finding the “**best” linear predictor** of Y in terms of X.

- This means that the <u>expected</u> Mean Square Error

$$\mathbb{E}[(Y - (b_0 + b_1 X))^2]$$

  is minimized.

- **Approach:** This can be done (using calculus!) to minimize the function MSE(b_0, b_1) as a function of the two parameters b_0 and b_1.
- **Idea:** Expand this as a function (with numbers!) of b_0 and b_1. Then the minimum will occur when the derivatives (w.r.t. b_0 and b_1) are zero!

  - $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 x))^2]}{\partial b_0} = \frac{\partial \mathbb{E}\left[Y^2 - 2(b_0 + b_1 x)Y + (b_0 + b_1 x)^2\right]}{\partial b_0}$

$$= -2\mathbb{E}[Y] + 2b_0 + 2b_1\mathbb{E}[X] = 0.$$

$$\boxed{\Rightarrow\ \mathbb{E}[Y] = b_0 + b_1\mathbb{E}[X]}$$

What is the best linear predictor?

- Similarly we can compute the derivative with respect to b_1:

What is the best linear predictor?

- Similarly we can compute the derivative with respect to b_1:

    - $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 X))^2]}{\partial b_1} =$

What is the best linear predictor?

- Similarly we can compute the derivative with respect to b_1:

- $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 X))^2]}{\partial b_1} = \frac{\partial \mathbb{E}\left[Y^2 - 2(b_0 + b_1 X)Y + (b_0 + b_1 X)^2\right]}{\partial b_1} = \ldots = 0$

***Footnote:** See 2019 Lecture Notes by C. Shalizi on "The Truth About Linear Regression" available at https://www.stat.cmu.edu/~cshalizi/TALR

What is the best linear predictor?

- Similarly we can compute the derivative with respect to b_1:

- $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 X))^2]}{\partial b_1} = \frac{\partial \mathbb{E}\left[Y^2 - 2(b_0 + b_1 X)Y + (b_0 + b_1 X)^2\right]}{\partial b_1} = \ldots = 0$

$$\boxed{\Rightarrow \; b_1 = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}}$$

[Reference:TALR* pp14-15].

***Footnote:** See 2019 Lecture Notes by C. Shalizi on "The Truth About Linear Regression" available at https://www.stat.cmu.edu/~cshalizi/TALR

What is the best linear predictor?

- Similarly we can compute the derivative with respect to b_1:

- $\frac{\partial \mathbb{E}[(Y-(b_0+b_1X))^2]}{\partial b_1} = \frac{\partial \mathbb{E}\left[Y^2 - 2\left(b_0+b_1X\right)Y + (b_0+b_1X)^2\right]}{\partial b_1}$ = … = 0

$$\boxed{\Rightarrow\ b_1 = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}}$$

[Reference:TALR* pp14-15].

Summary:

***Footnote:** See 2019 Lecture Notes by C. Shalizi on "The Truth About Linear Regression" available at https://www.stat.cmu.edu/~cshalizi/TALR

What is the best linear predictor?

- Similarly we can compute the derivative with respect to b_1:

- $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 X))^2]}{\partial b_1} = \frac{\partial \mathbb{E}\left[Y^2 - 2(b_0 + b_1 X)Y + (b_0 + b_1 X)^2\right]}{\partial b_1} = \ldots = 0$

$$\boxed{\Rightarrow \ b_1 = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}}$$

[Reference:TALR* pp14-15].

Summary:

- $b_1 = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}$.

***Footnote:** See 2019 Lecture Notes by C. Shalizi on "The Truth About Linear Regression" available at https://www.stat.cmu.edu/~cshalizi/TALR

What is the best linear predictor?

- Similarly we can compute the derivative with respect to b_1:

- $\frac{\partial \mathbb{E}[(Y - (b_0 + b_1 X))^2]}{\partial b_1} = \frac{\partial \mathbb{E}\left[Y^2 - 2(b_0 + b_1 X)Y + (b_0 + b_1 X)^2\right]}{\partial b_1} = \ldots = 0$

$$\boxed{\Rightarrow \; b_1 = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}}$$

[Reference:TALR* pp14-15].

Summary:

- $b_1 = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}$.
- b_0 ensures that the point $(\mathbb{E}[X], \mathbb{E}[Y])$ is on the line.

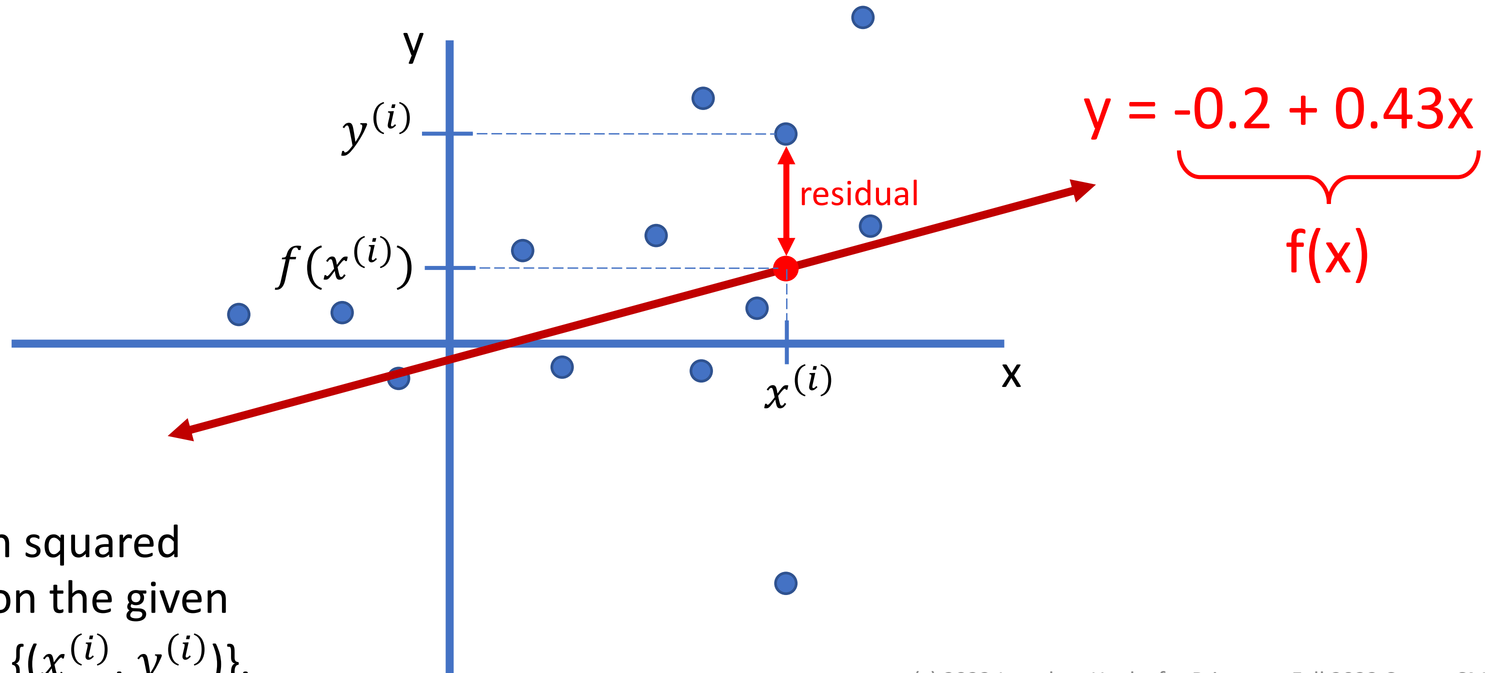
***Footnote:** See 2019 Lecture Notes by C. Shalizi on "The Truth About Linear Regression" available at https://www.stat.cmu.edu/~cshalizi/TALR

Residuals and the “best” linear model

- The best line minimizes the (mean squared) error, i.e. minimizes the number

Mean Squared Error (MSE) :=

$$\frac{1}{n}\left[\left(y^{(1)} - f\left(x^{(1)}\right)\right)^2 + \cdots + \left(y^{(n)} - f\left(x^{(n)}\right)\right)^2\right]$$



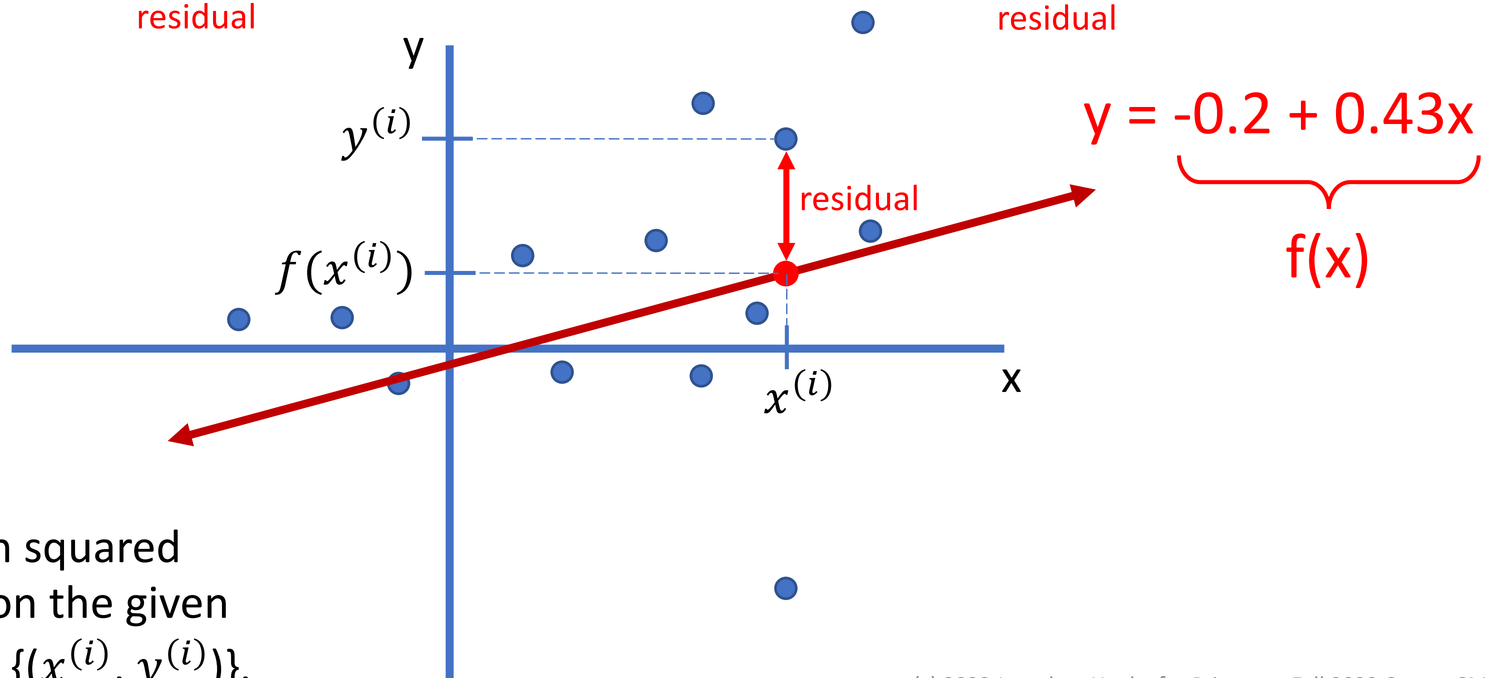
Note: The mean squared error depends on the given dataset of pairs $\{(x^{(i)}, y^{(i)})\}$.

Residuals and the "best" linear model

- The best line minimizes the (mean squared) error, i.e. minimizes the number

Mean Squared Error (MSE) :=

$$\frac{1}{n}\left[\Big(\underbrace{y^{(1)} - f\big(x^{(1)}\big)}_{\text{residual}}\Big)^2 + \cdots + \Big(\underbrace{y^{(n)} - f\big(x^{(n)}\big)}_{\text{residual}}\Big)^2\right]$$



Note: The mean squared error depends on the given dataset of pairs $\{(x^{(i)}, y^{(i)})\}$.

Understanding Residuals

- **Residuals** are the differences between the actual values and the predicted values. The MSE is the sum of the squares of the residuals.

- The (MSE) "best" linear regression line will have residuals with mean zero. This is true for the residual random variable, but also for the statistic measuring the residual on actual data (i.e. the computed residuals for any given x, and over all x.)

- I.e. $Y = f(X) + r(X)$

$$\Rightarrow \mathbb{E}[Y] = \mathbb{E}[f(X)] + \mathbb{E}[r(X)],$$

but $(\mathbb{E}[X], \mathbb{E}[Y])$ is already on the linear regression line,

so $\mathbb{E}[r(X)] = 0$.

Assumptions of Linear Regression

Assumptions of Linear Regression

1. X and Y are random variables.

Assumptions of Linear Regression

1. X and Y are random variables.
2. X and Y have some underlying relationship
   $\mathrm{Y} = f(X) + \varepsilon(X)$.

Assumptions of Linear Regression

1. X and Y are random variables.
2. X and Y have some underlying relationship $\mathrm{Y} = f(X) + \varepsilon(X)$.
3. The noise $\varepsilon(X)$ has mean zero for any given X.

Relation to Correlation

Relation to Correlation

- You may have already used the term “correlation” between two random variables, which expresses how linearly related they are (on average).

Relation to Correlation

- You may have already used the term “correlation” between two random variables, which expresses how linearly related they are (on average).

- In our review, we recalled that the (Pearson) **correlation coefficient** is closely related to the covariance of the two random variables, defined as

$$\rho = \frac{\mathrm{Cov}[X,Y]}{\mathrm{StdDev}[X] \cdot \mathrm{StdDev}[Y]}.$$

Relation to Correlation

- You may have already used the term "correlation" between two random variables, which expresses how linearly related they are (on average).

- In our review, we recalled that the (Pearson) **correlation coefficient** is closely related to the covariance of the two random variables, defined as

$$\rho = \frac{\mathrm{Cov}[X,Y]}{\mathrm{StdDev}[X] \cdot \mathrm{StdDev}[Y]}.$$

- This is closely related to the slope of the (MSE) best line

$$b_1 = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}.$$

Mid-class Break!

- Stretch your legs / get a snack break!

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y := a_0 \; + \; a_1 X \; + \varepsilon$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y := a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y := a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε “epsilon” have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]}$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y := a_0 \ + \ a_1 X \ + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[X]} = \frac{\mathrm{Cov}[X, a_0 + a_1 X + \varepsilon\,]}{\mathrm{Var}[X]}$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y := a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon]}{\text{Var}[X]}$$

$$= \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon]}{\text{Var}[X]}$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y := a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon\,]}{\text{Var}[X]}$$

$$= \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon\,]}{\text{Var}[X]}$$

$$= \frac{0 + a_1\,\text{Var}[X] + 0}{\text{Var}[X]} = a_1$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y := a_0 \ + \ a_1 X \ + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon\ ]}{\text{Var}[X]}$$

$$= \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon\ ]}{\text{Var}[X]}$$

$$= \frac{0 + a_1 \, \text{Var}[X]\ + 0]}{\text{Var}[X]} = \ a_1$$

so linear regression correctly recovers b_1 as the slope a_1 of the underlying relationship!

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y \; = a_0 \; + \; a_1 X \; + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon\,]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon\,]}{\text{Var}[X]}$$

$$= \frac{0 + a_1\,\text{Var}[X] \; + 0]}{\text{Var}[X]} = \; a_1$$

so linear regression correctly recovers b_1 as the slope a_1 of the underlying relationship!

Also we can understand b_0 by computing the expected value

$$\mathbb{E}[Y]$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y = a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon]}{\text{Var}[X]}$$

$$= \frac{0 + a_1 \text{Var}[X] + 0]}{\text{Var}[X]} = a_1$$

so linear regression correctly recovers b_1 as the slope a_1 of the underlying relationship!

Also we can understand b_0 by computing the expected value

$$\mathbb{E}[Y] = \mathbb{E}[a_0 + a_1 X + \varepsilon]$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y = a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon\,]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon\,]}{\text{Var}[X]}$$

$$= \frac{0 + a_1 \text{Var}[X] \; + 0]}{\text{Var}[X]} = a_1$$

so linear regression correctly recovers b_1 as the slope a_1 of the underlying relationship!

Also we can understand b_0 by computing the expected value

$$\mathbb{E}[Y] = \mathbb{E}[a_0 + a_1 X + \varepsilon] = \mathbb{E}[a_0] + \mathbb{E}[a_1 X] + \mathbb{E}[\varepsilon]$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y = a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon]}{\text{Var}[X]}$$

$$= \frac{0 + a_1 \text{Var}[X] + 0]}{\text{Var}[X]} = a_1$$

so linear regression correctly recovers b_1 as the slope a_1 of the underlying relationship!

Also we can understand b_0 by computing the expected value

$$\mathbb{E}[Y] = \mathbb{E}[a_0 + a_1 X + \varepsilon] = \mathbb{E}[a_0] + \mathbb{E}[a_1 X] + \mathbb{E}[\varepsilon] = a_0 + a_1 \mathbb{E}[X] + 0,$$

Special Case: Linearly related RVs

Consider the special case where the random variable Y is given by

$$Y = a_0 + a_1 X + \varepsilon$$

for some (non-random) numbers a_0 and a_1, and the (RV) noise ε "epsilon" have mean zero (i.e. $\mathbb{E}[\varepsilon] = 0$) and are uncorrelated with the random variable X.

Then we see that linear regression gives:

$$b_1 = \frac{\text{Cov}[X,Y]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0 + a_1 X + \varepsilon]}{\text{Var}[X]} = \frac{\text{Cov}[X, a_0] + \text{Cov}[X, a_1 X] + \text{Cov}[X, \varepsilon]}{\text{Var}[X]}$$

$$= \frac{0 + a_1 \text{Var}[X] + 0]}{\text{Var}[X]} = a_1$$

so linear regression correctly recovers b_1 as the slope a_1 of the underlying relationship!

Also we can understand b_0 by computing the expected value

$$\mathbb{E}[Y] = \mathbb{E}[a_0 + a_1 X + \varepsilon] = \mathbb{E}[a_0] + \mathbb{E}[a_1 X] + \mathbb{E}[\varepsilon] = a_0 + a_1 \mathbb{E}[X] + 0,$$

so since $b_1 = a_1$, we see that the given coefficient a_0 also agrees with the coefficient b_0 appearing in the linear regression!

Remarks on Linear Regression for RVs

Remarks on Linear Regression for RVs

1. No assumptions about the actual linearity of the true relationship between the random variables X and Y to find the (MSE) “best line” relating them. This relationship <u>may or may not</u> be linear.

Remarks on Linear Regression for RVs

1. No assumptions about the actual linearity of the true relationship between the random variables X and Y to find the (MSE) “best line” relating them. This relationship <u>may or may not</u> be linear.

2. Linear regressions <u>may or may not</u> give an overall small (MSE) error.

Remarks on Linear Regression for RVs

1. No assumptions about the actual linearity of the true relationship between the random variables X and Y to find the (MSE) “best line” relating them. This relationship may or may not be linear.

2. Linear regressions may or may not give an overall small (MSE) error.

3. The slope is closely related to the correlation between X and Y.

Remarks on Linear Regression for RVs

1. No assumptions about the actual linearity of the true relationship between the random variables X and Y to find the (MSE) "best line" relating them. This relationship <u>may or may not</u> be linear.

2. Linear regressions <u>may or may not</u> give an overall small (MSE) error.

3. The slope is closely related to the correlation between X and Y.

4. In general, the linear regression line will depend on the distribution of X, even if the relationship between Y and X is unchanged! It also doesn't make sense in general to talk about the linear regression line for Y without specifying an underlying probability distribution for X, which in general is outside of our control (i.e. comes from an independent variable that we're observing in the real world)!

Remarks on Linear Regression for RVs

1. No assumptions about the actual linearity of the true relationship between the random variables X and Y to find the (MSE) “best line” relating them. This relationship <u>may or may not</u> be linear.

2. Linear regressions <u>may or may not</u> give an overall small (MSE) error.

3. The slope is closely related to the correlation between X and Y.

4. In general, the linear regression line will depend on the distribution of X, even if the relationship between Y and X is unchanged! It also doesn’t make sense to talk about the linear regression line for Y in general without specifying an underlying probability distribution for X, which in general is outside of our control (i.e. comes from an independent variable that we’re observing in the real world)!

5. In the special case where Y is linearly related to X (plus some mean zero noise not correlated to X), we see that Linear Regression recovers the coefficients a_0 and a_1 of linear relationship. In this case the underlying probability distribution of X does not play a role.