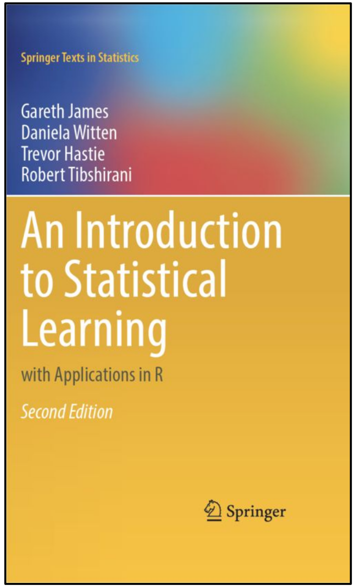
Additional Metrics for Regression

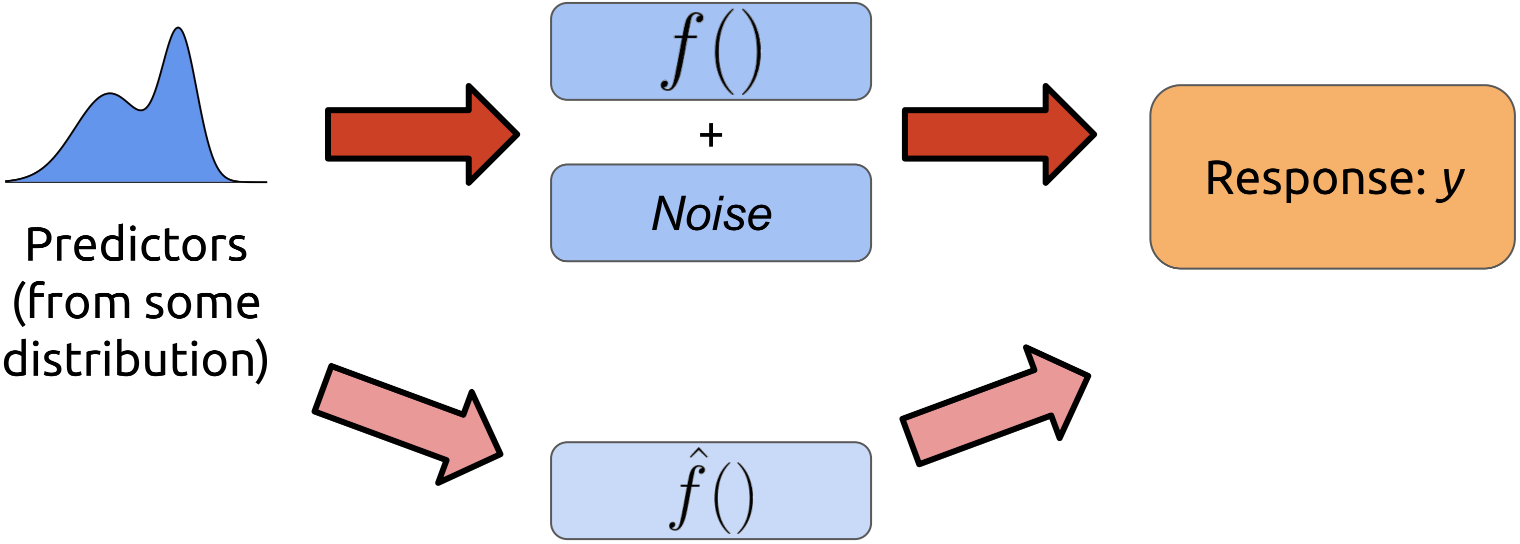
Recommended Reading:

An Introduction to Statistical Learning, Section 2.2.1

Download it for free here:
https://www.statlearning.com/



Supervised Learning - Goals



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

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.

Very 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.

Measuring Generalization Data - How

If we only care about how well our model performs on unseen data, how do we measure that?

Measuring Generalization Data - How

If we only care about how well our model performs on unseen data, how do we measure that?

We can’t - it’s unseen!

Measuring Generalization Data - How

If we only care about how well our model performs on unseen data, how do we measure that?

We can’t - it’s unseen!

But, we can *estimate* it.

Measuring Generalization Data - How

The most simple way to estimate generalization error is through employing a train/test split.

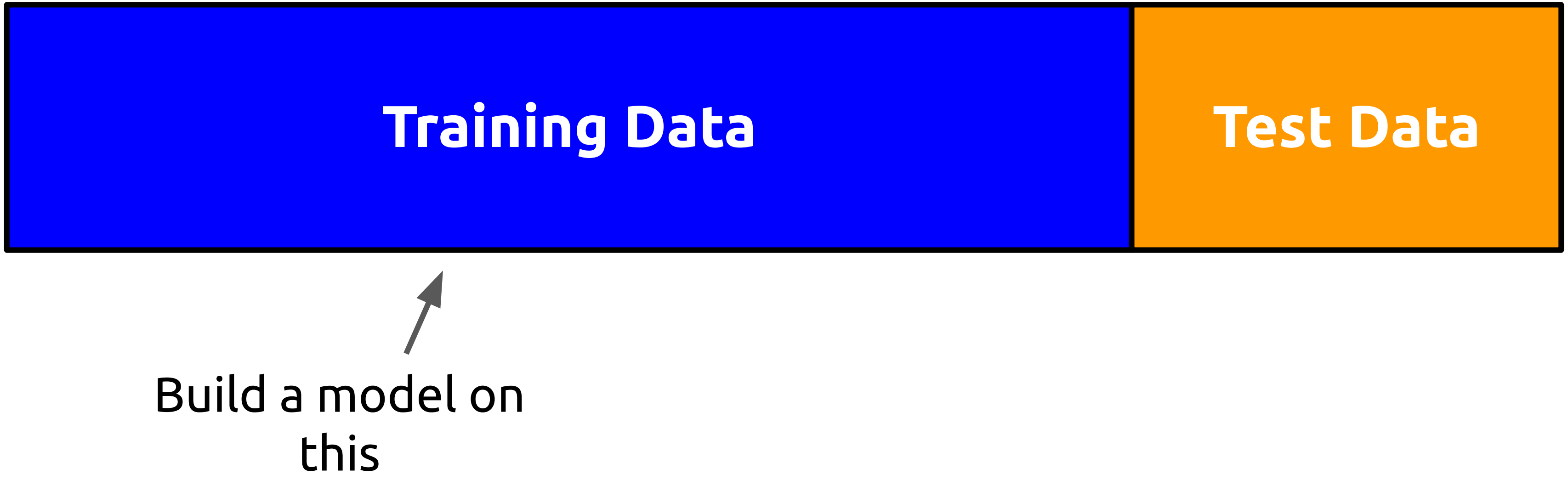
Full Dataset

Measuring Generalization Data - How

The most simple way to estimate generalization error is through employing a train/test split.

Measuring Generalization Data - How

The most simple way to estimate generalization error is through employing a train/test split.



Measuring Generalization Data - How

The most simple way to estimate generalization error is through employing a train/test split.

Build a model on this

Estimate generalization error by seeing how it performs on this.

Measuring Generalization Data - How

The most simple way to estimate generalization error is through employing a train/test split.

How well does this work? Let’s jump into a notebook to see.

Regression Metrics

Some common metrics for judging a regression model are

- (Root) Mean Squared Error
- Mean Absolute Error
- Mean Absolute Percentage Error
- R^2

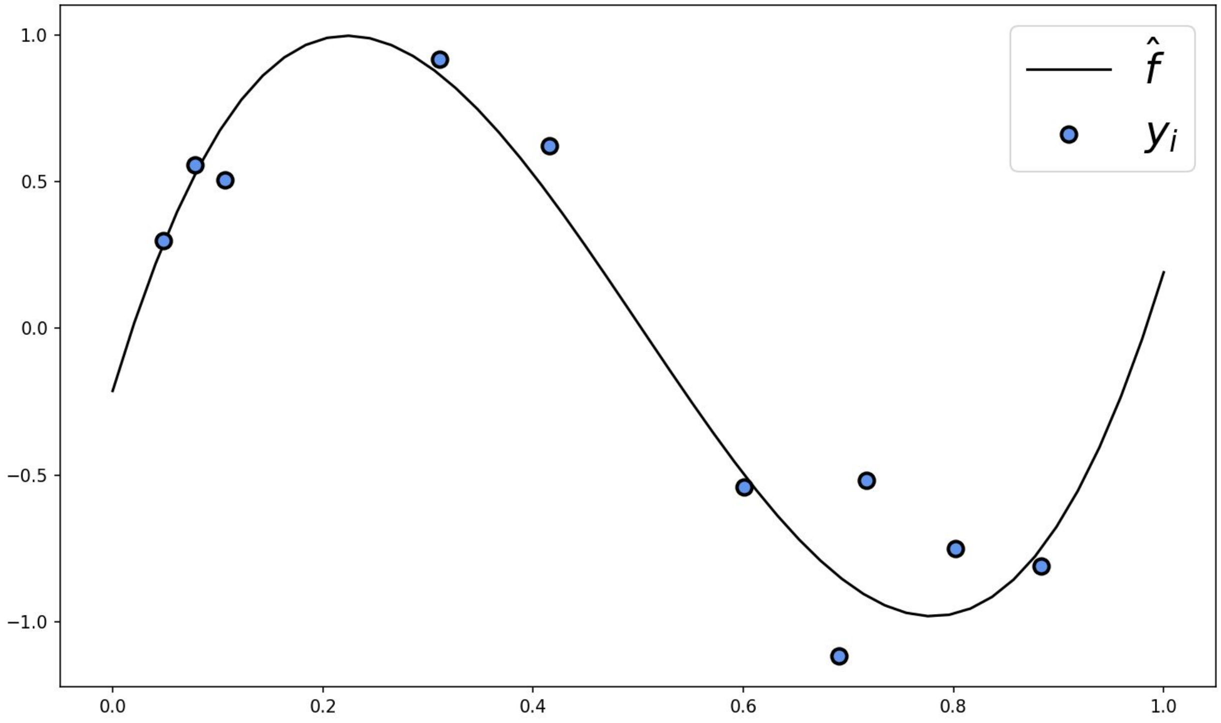
Residuals

Given a dataset $\{(x_1, y_1), (x_2, y_2),\ldots,(x_n, y_n)\}$ and a predictor function $\hat{f}$ the residuals are given by:

$$\mathrm{residual}_i = \hat{f}(x_i) - y_i$$

Residuals

$$\text{residual}_i = \hat{f}(x_i) - y_i$$



Mean Squared Error

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

Mean Squared Error

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

$$\text{residual}_i = \hat{f}(x_i) - y_i$$

Mean Squared Error

When fitting a linear regression model, the usual way it is done is by minimizing the MSE on the *training* data.

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

Mean Squared Error

When fitting a linear regression model, the usual way it is done is by minimizing the MSE on the *training* data.

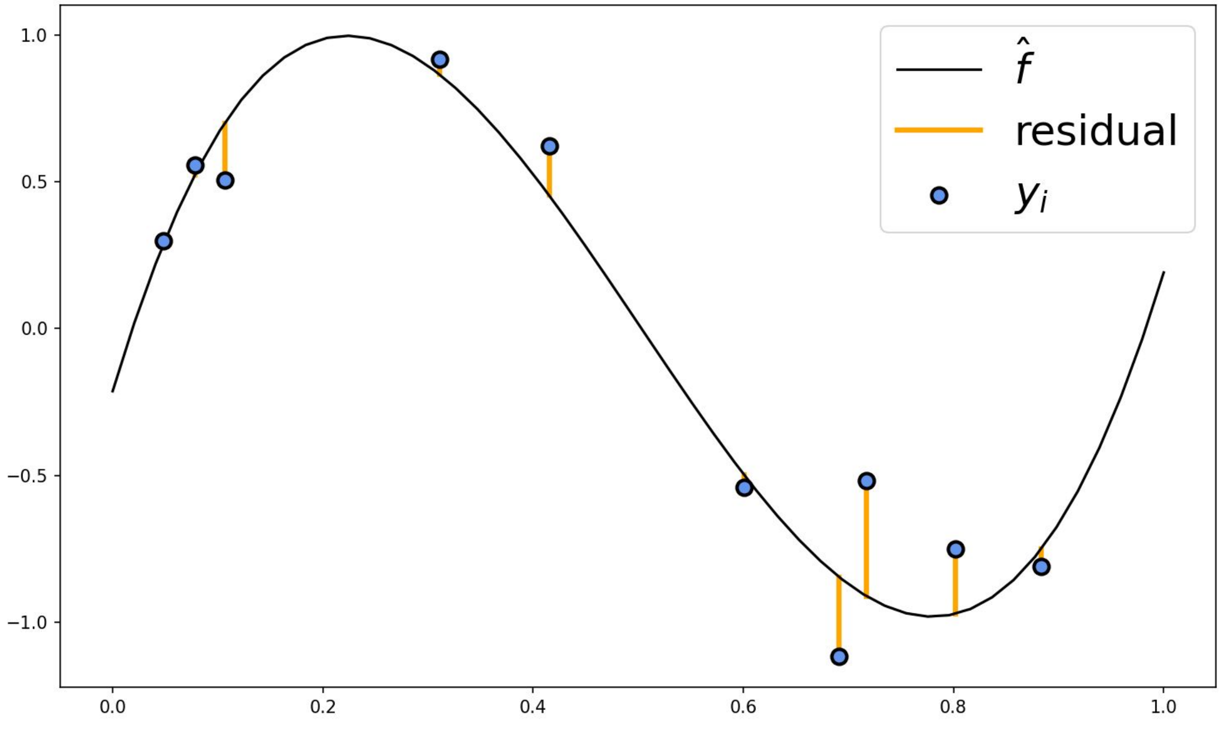
$$MSE = \frac{\sum_{i=1}^{n}(\hat{f}(x_i) - y_i)^2}{n}$$

Downsides: It is measured in square units.

Eg. if predicting price, MSE will be in squared dollars.

Outlier values can be highly influential on MSE.

Mean Squared Error



MSE = 0.035

Root Mean Squared Error

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

Root Mean Squared Error

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

Pros: It is measured in the same units as the target.

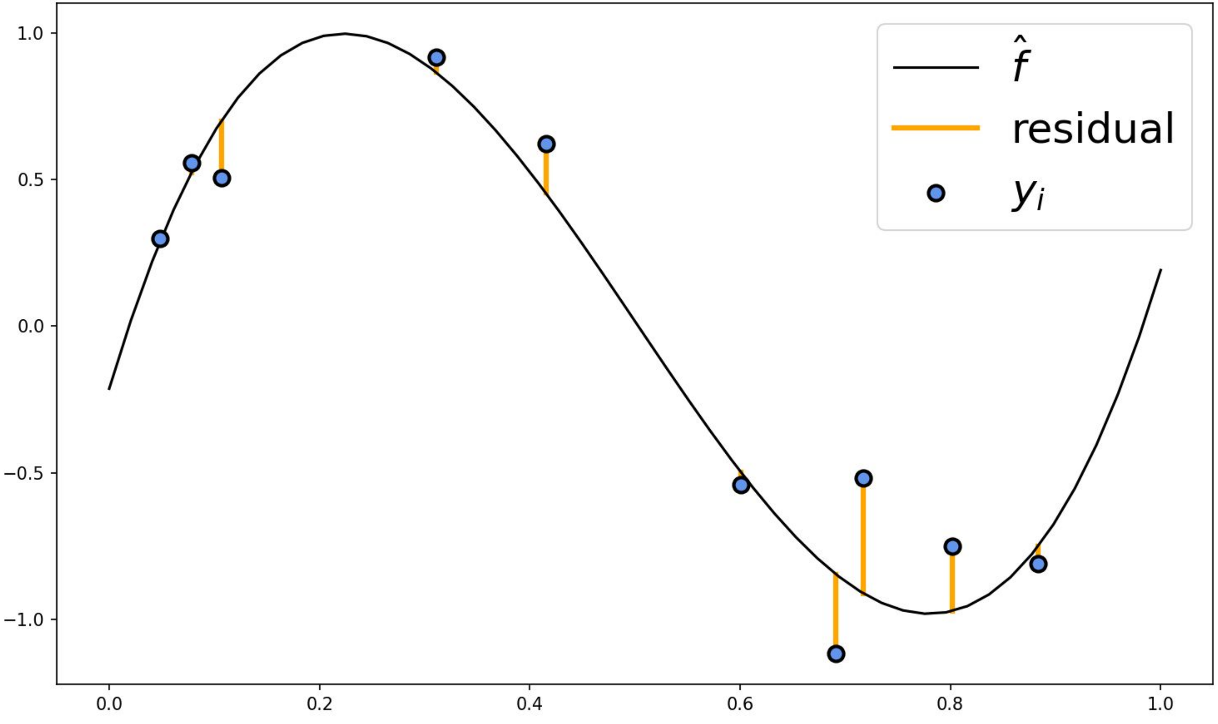
Root Mean Squared Error

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

Pros: It is measured in the same units as the target.

Cons: It is a bit difficult to directly understand what it is measuring.

Root Mean Squared Error



RMSE = 0.188

Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^{n} |\hat{f}(x_i) - y_i|}{n}$$

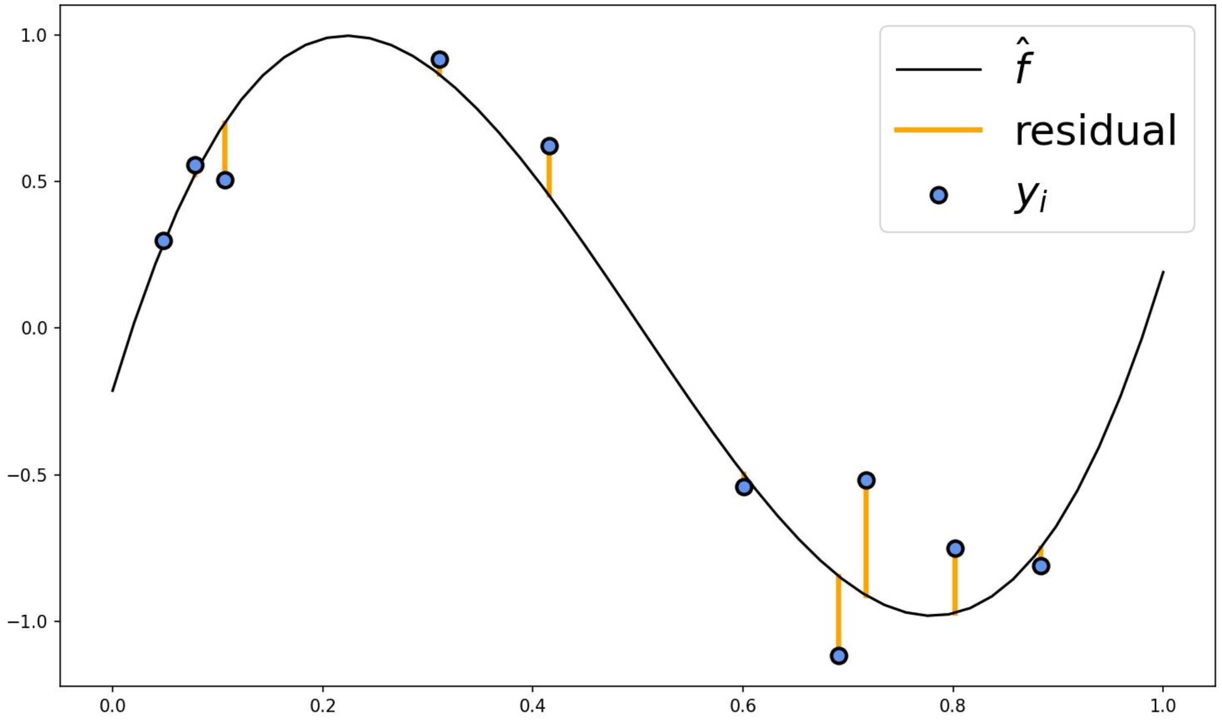
Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^{n} |\hat{f}(x_i) - y_i|}{n}$$

Pros: Very easy to interpret - it is exactly the average error.

Cons: Don’t have a nice way to minimize it on the training data (like we can with MSE).

Mean Absolute Error



MAE = 0.145

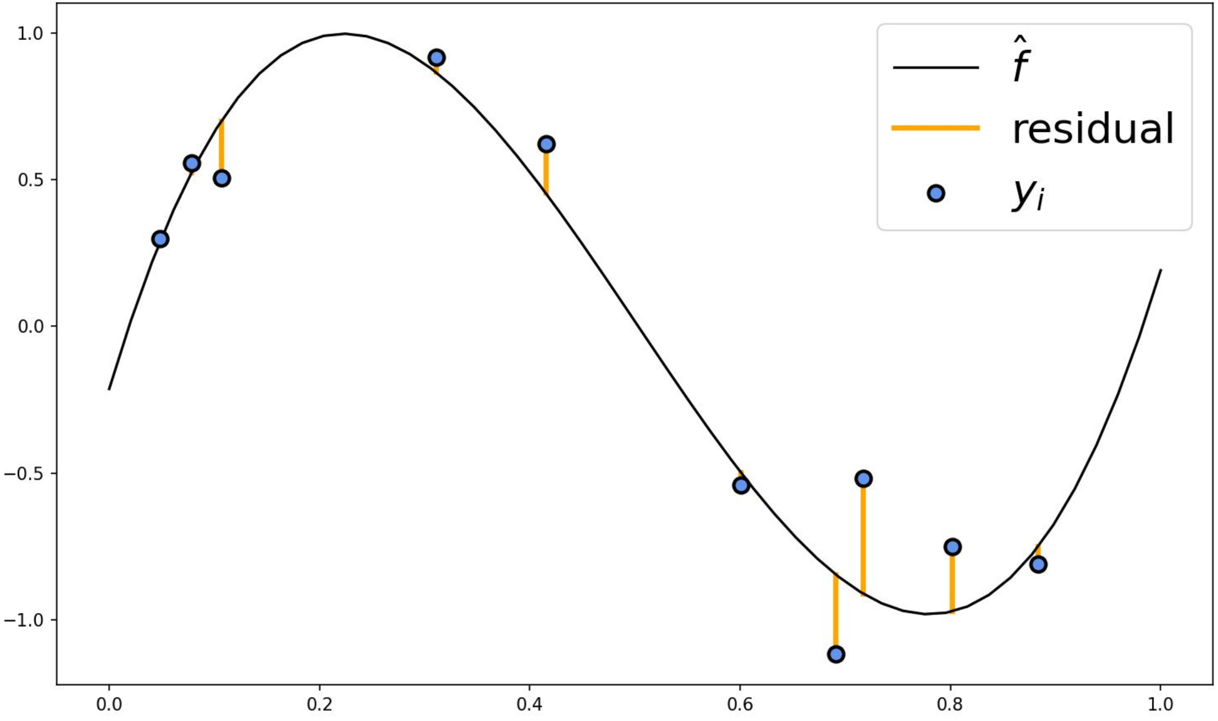
Mean Absolute Percentage Error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{f}(x_i) - y_i}{y_i}\right|$$

Pros: Normalizes the error by the size of the target. If you have a wide range of target values, this can be more informative than just calculating absolute errors.

Cons: Doesn’t work if any of the targets are 0.

Mean Absolute Percentage Error



MAPE = 0.226

R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

$$TSS = \sum_{i=1}^{n} (\bar{y} - y_i)^2$$

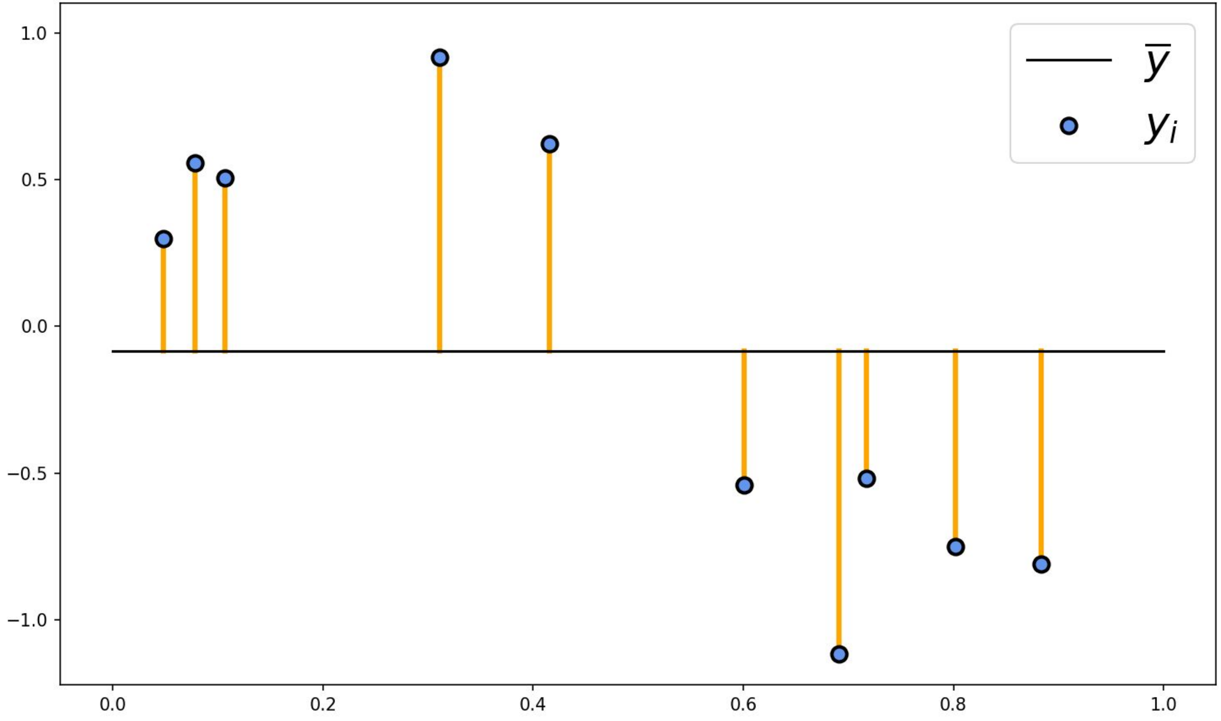
R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

$$TSS = \sum_{i=1}^{n} (\boxed{\bar{y}} - y_i)^2$$

the average target value

Total Sum of Squares (TSS)



TSS = 4.845

R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

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

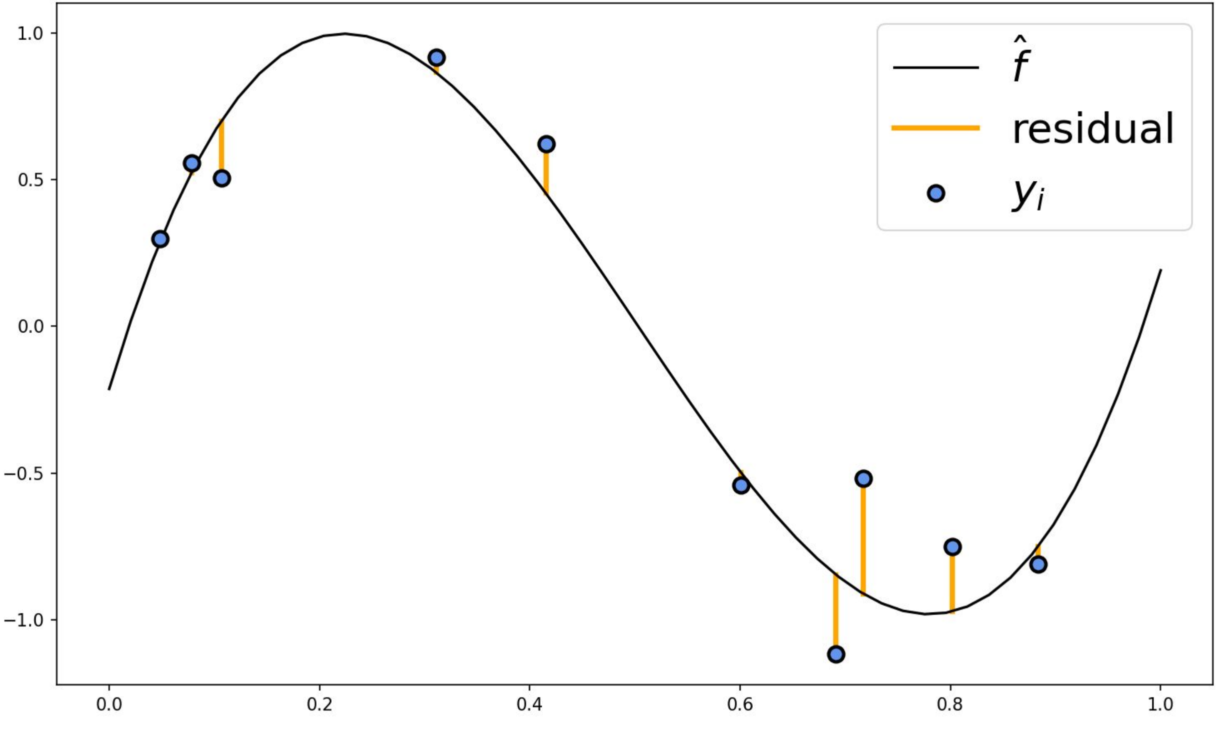
R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

$$RSS = \sum_{i=1}^{n} \boxed{(\hat{f}(x_i) - y_i)}^2$$

$$\text{residual}_i = \hat{f}(x_i) - y_i$$

Residual Sum of Squares (RSS)



RSS = 0.352

R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

$$R^2 = \frac{TSS - RSS}{TSS}$$

R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

$$R^2 = \frac{TSS - RSS}{TSS}$$

The best possible R^2 of 1 occurs when RSS = 0.

R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

$$R^2 = \frac{TSS - RSS}{TSS}$$

The best possible R^2 of 1 occurs when RSS = 0.

The simplest model (predict the average target value) has an R^2 value of 0.

R^2

Compares the **residual sum of squares (RSS)** to the **total sum of squares (TSS)**.

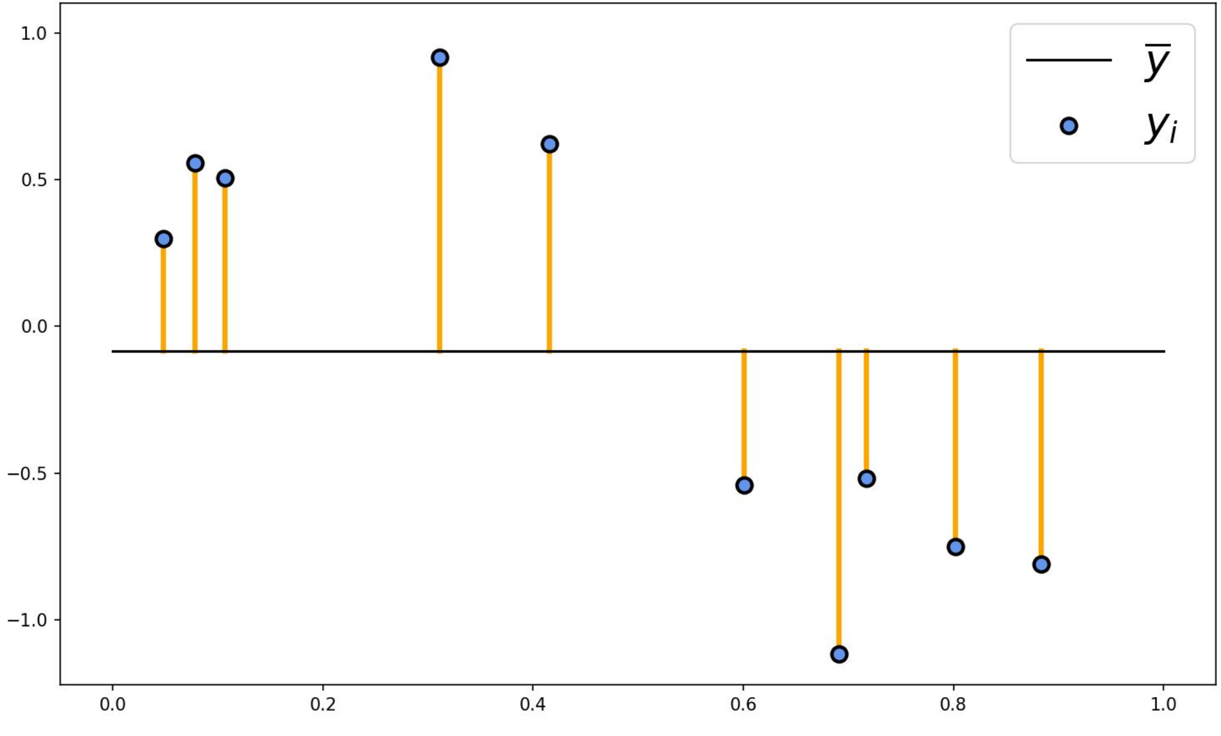
$$R^2 = \frac{TSS - RSS}{TSS}$$

The best possible R^2 of 1 occurs when RSS = 0.

The simplest model (predict the average target value) has an R^2 value of 0.

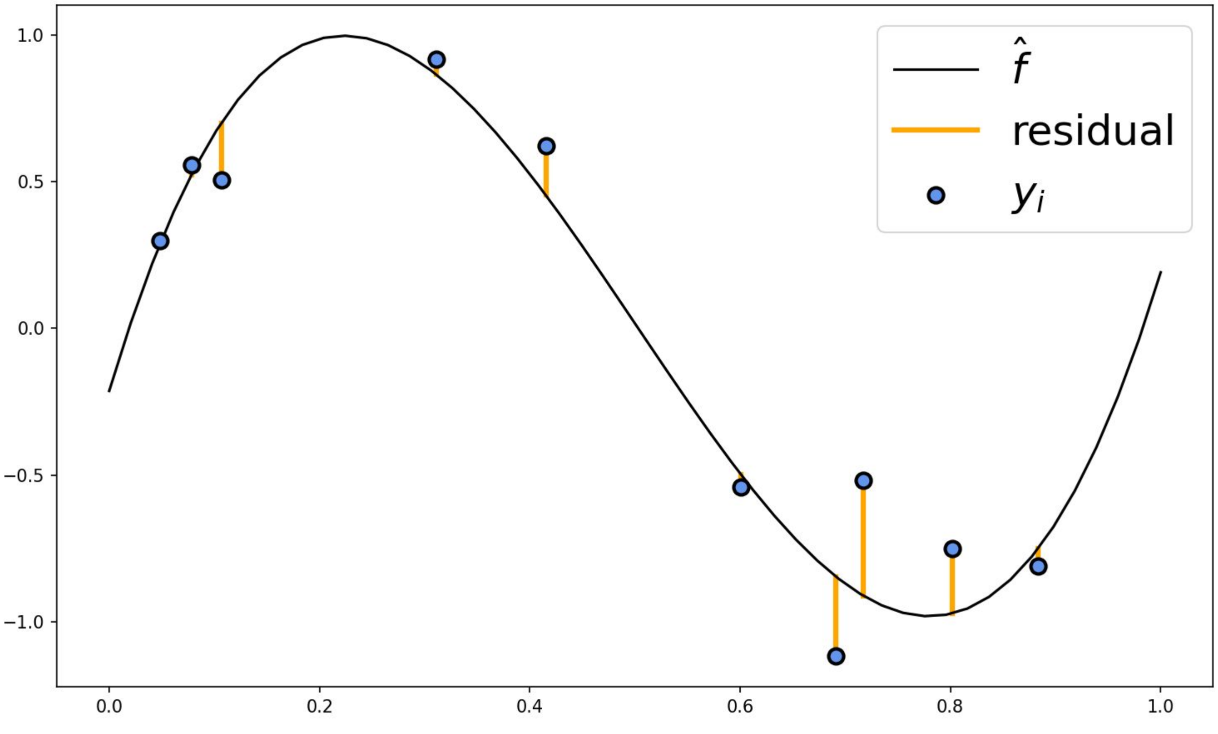
The range of R^2 is $(-\infty, 1]$.

Total Sum of Squares (TSS)



TSS = 4.845

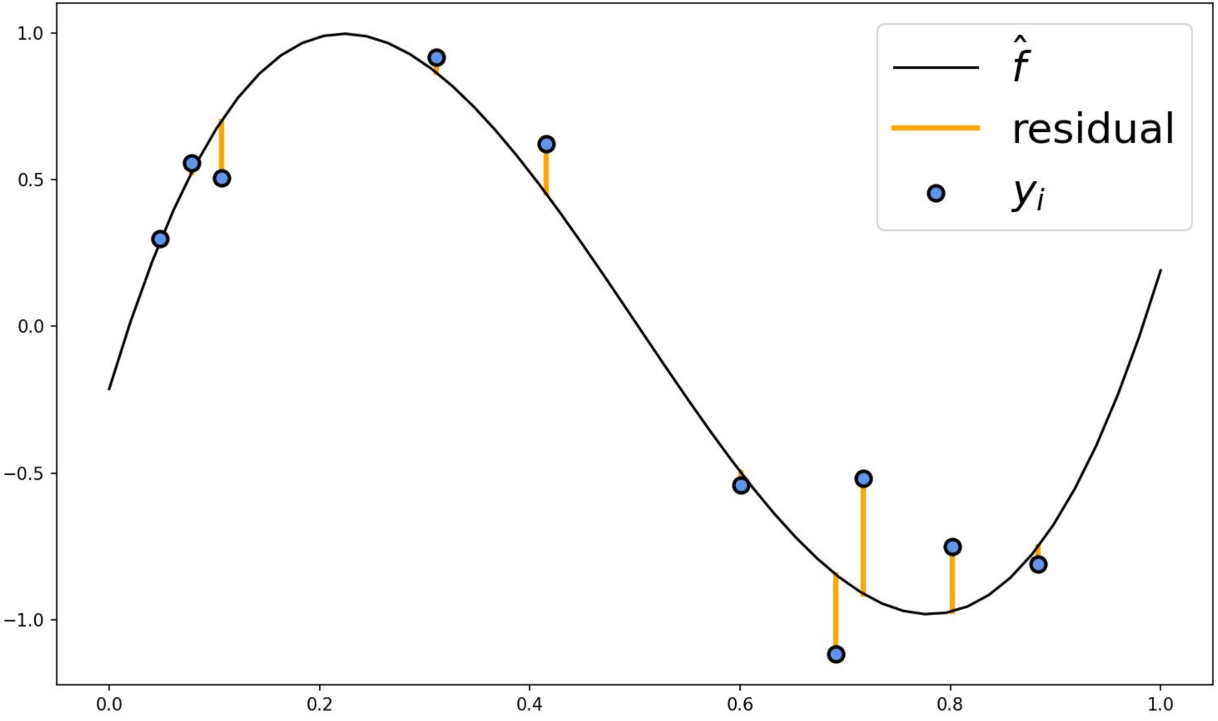
Residual Sum of Squares (RSS)



TSS = 4.845

RSS = 0.352

R^2



TSS = 4.845

RSS = 0.352

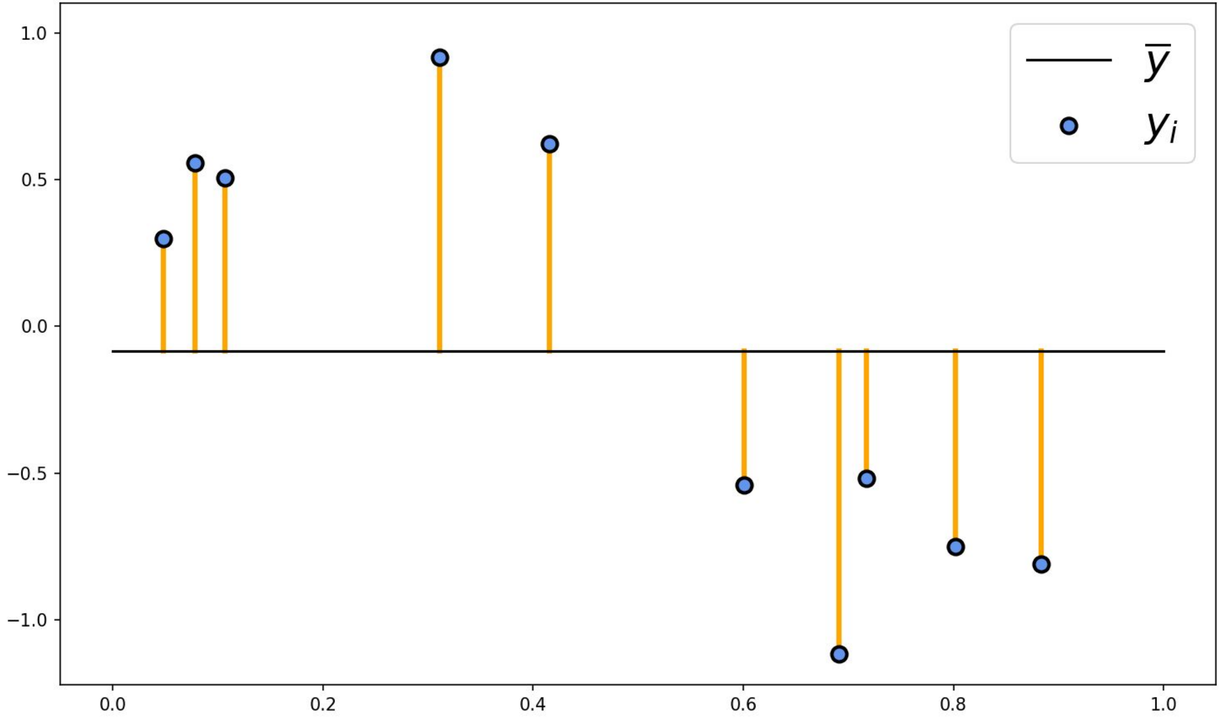
$$R^2 = \frac{4.845 - 0.352}{4.845}$$

$$= 0.931$$

R^2

What about a negative R^2 value?

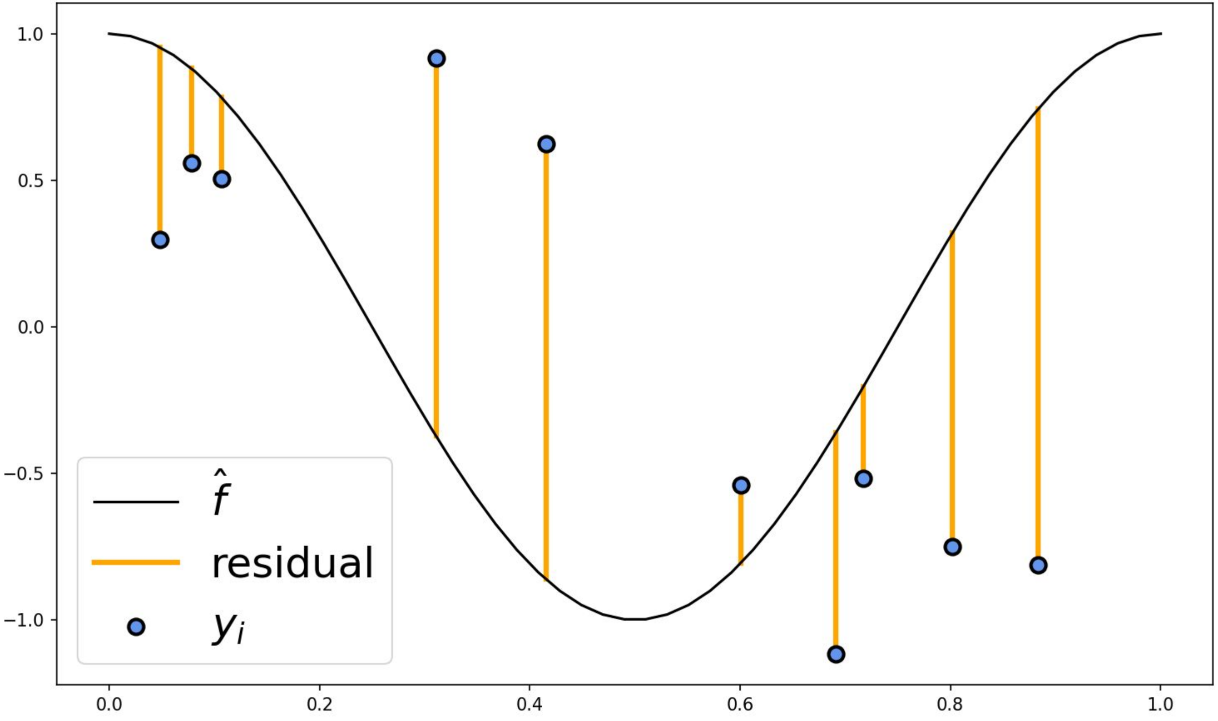
Total Sum of Squares (TSS)



TSS = 4.845

RSS = 0.352

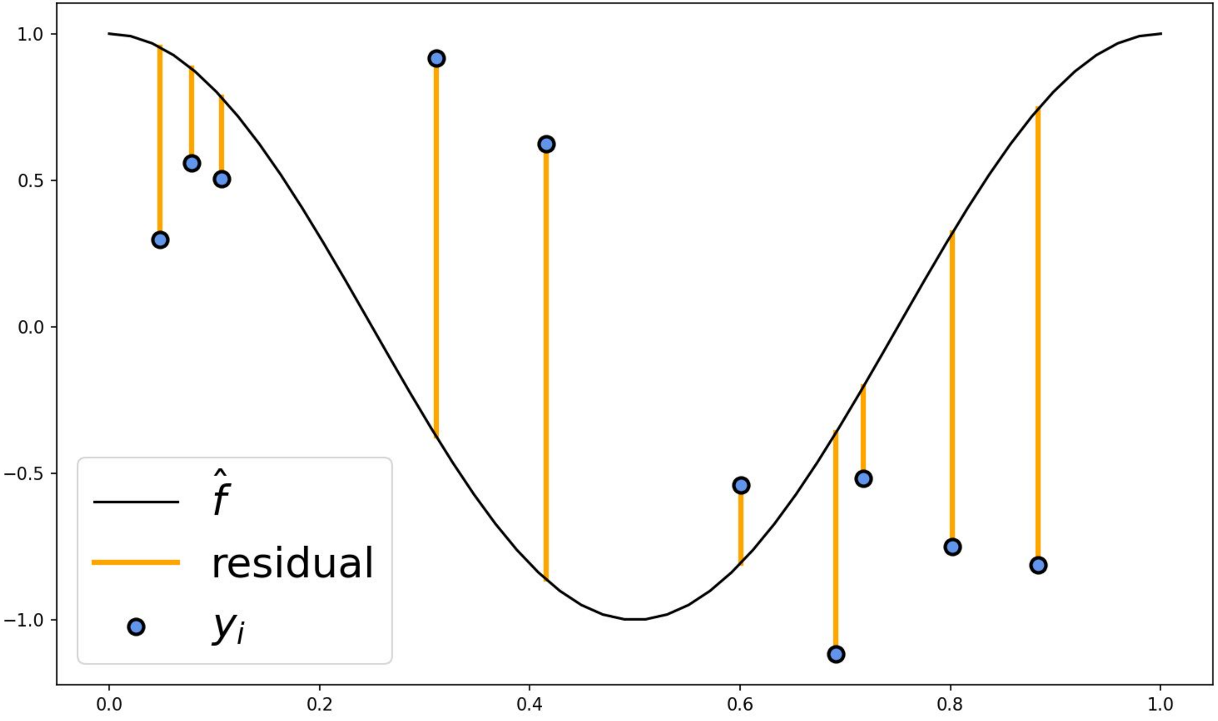
Residual Sum of Squares (RSS)



TSS = 4.845

RSS = 97.58

R^2



TSS = 4.845

RSS = 97.58

$$R^2 = \frac{4.845 - 97.58}{4.845}$$

$$= -19.140$$