How do we Evaluate?

Now we need to decide which model is the best.

How we do that?

THE FIRST THING IS :

We Evaluate our errors :

Based on the model ability to Figure out a new data THAT IS PREDETERMENT.

↓

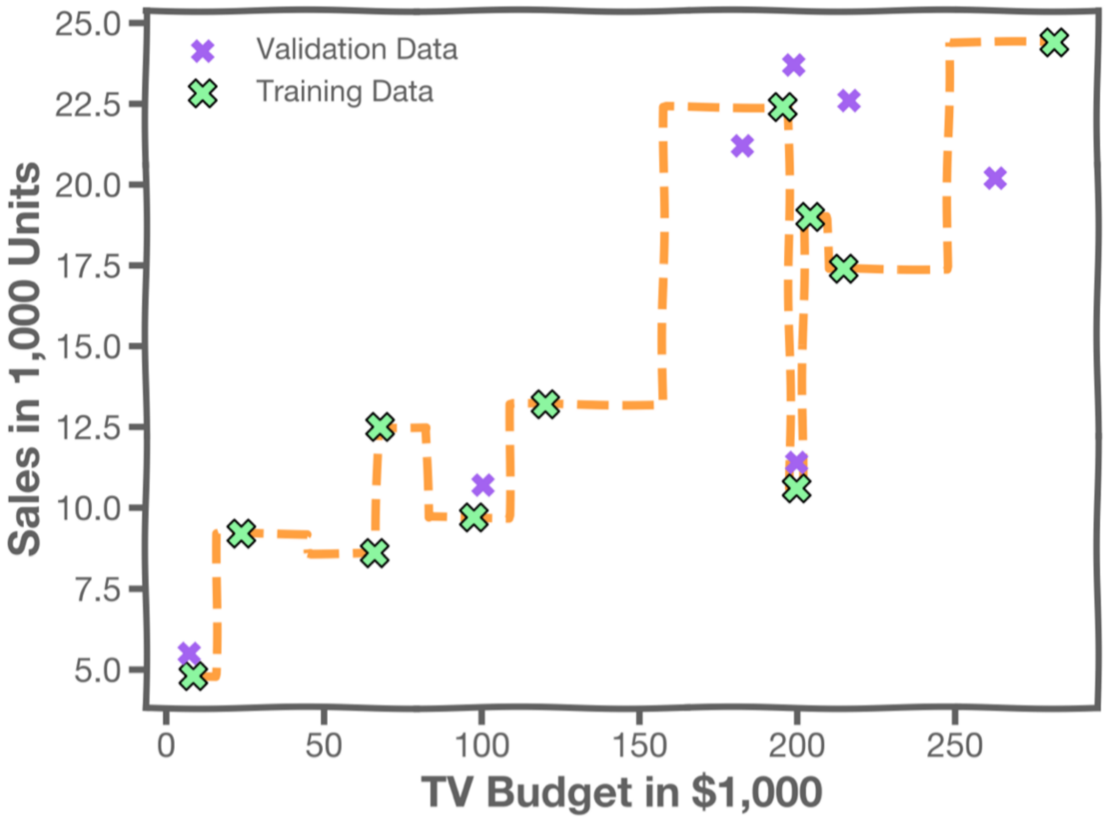
Than we can evaluate the model ability to predict a new data based on the % he actually able to pinpoint a new data.

And for our example, we are focusing on the statistical model with the highest % predictability

We start with a model of $k = 1$

We start with a set of data and randomly hide some of that data from our model. This is called a train-validation split. We use the visible part of the data (the training set) to estimate $\hat{y}$, and the hidden part (the validation set) to evaluate the model.

The one-neighbor model (k=1) used to make predictions $\hat{y}$ using the training set is shown on the plot. Now, we look at the data we have *not* used to make the model, the validation data shown as purple crosses.



The difference between the true value (the red cross) and the prediction is called the residual.

In this example we can see a clear Overfitting. The statistical model don't even try to predict and only focus on the data he knows.

Error Evaluation Continued

In order to quantify how well a model performs, we aggregate the errors across the data, and we call that the **loss**, **error**, or **cost function**. Cost usually refers to the total loss, while loss refers to a single training point.

A common loss function for quantitative outcomes is the **Mean Squared Error (MSE):**

Mean Squared Error (MSE)

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

Watch Out!

The MSE is by no means the only valid loss function, and it's not always the best one to use! Other choices for loss function include:

- Max Absolute Error
- Mean Absolute Error
- Mean Squared Error

The square **R**oot of the **M**ean of the **S**quared **E**rrors (**RMSE**) is also commonly used.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Other Kinds of Errors

Numerical error isn't the only kind you'll have to worry about. Sometimes the error is more fundamental. Sometimes we end up putting data - or the person the data represents - into the wrong category. Listen to Nabib as he talks about Type 1 and Type 2 errors.

In Depth Explanations using Algo-trading logic

Error Evaluation Metrics

The "cost" of being wrong

① The Core Concepts: Residuals

* Math definition: The difference between the actual value (y) and the predicted value ($\bar{y}$)

  Formula: $e = y - \bar{y}$

* Trading logic: This is your PNL discrepancy per trade.

  Example: Gold price was 2050, Algo predicted 2040 ->

  The Residual = 10, Algo missed the target by 10 $.

② Max Absolute Error:

* Math Definition: The single largest error in the entire data set.

* Formula : $\max(|y_i - \bar{y}_i|)$
* Trading logic : "The Account Blower"
* Metric focus : This metric tells you the worst mistake your bot made in the backtest
* Why it's important ! : Even if the error is low BUT there are error spikes ->
One Hugh error/s (Max Error), can trigger a margin call.

* Help pinpointing error spikes -> for fixing them afterwards by changing the algo...

③ Mean Absolute Error :

Math Definition : The average of the absolute differences. We take the absolute value |...| because we don't want positive & negative values to cancel each other out.

* Formula : $\frac{1}{n} \sum |y_i - \bar{y}_i|$
* Trading logic : True overall error average.
* why it matters ! : It gives us a realistic

expectation of the realistic behavior of the bot in a normal day.

④ Mean Squered Error (MSE):

* Formula $\frac{1}{n}\sum(y_i - \hat{y}_i)^2$

* Because the square, big mistakes become a hugh number.

* Error of 2 become 4
* Error of 10 become 100

* Why it matters: It forces you to focus on the big outliers and fix them (change the algo accordingly).

⑤ RMSE Root Mean Squered Error:

* Formula $RMSE = \sqrt{MSE}$

* It is the standard deviation of your prediction errors:
It tells you to expect the price to be $\pm\sqrt{MSE}$

Model Comparison

Now we have a way to measure the Error of of the models by doing model comparison: