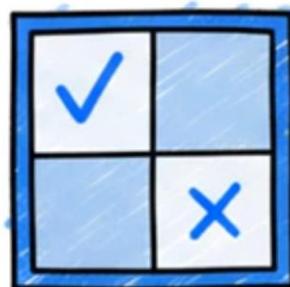
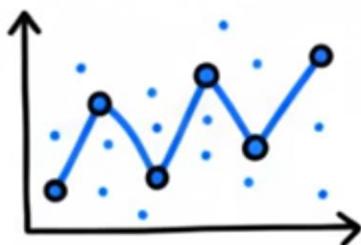
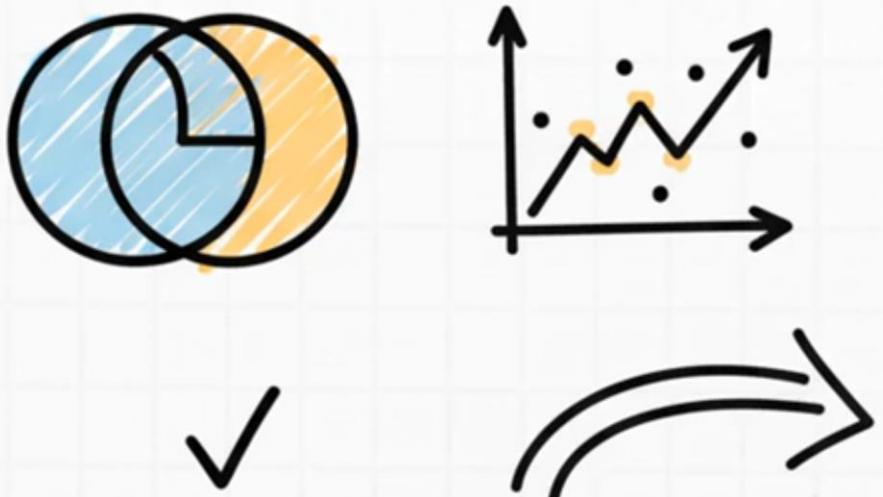


# Evaluating ML Models

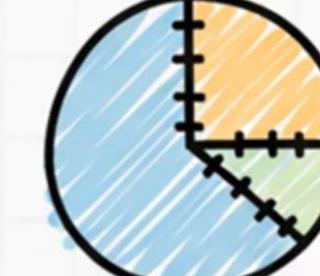
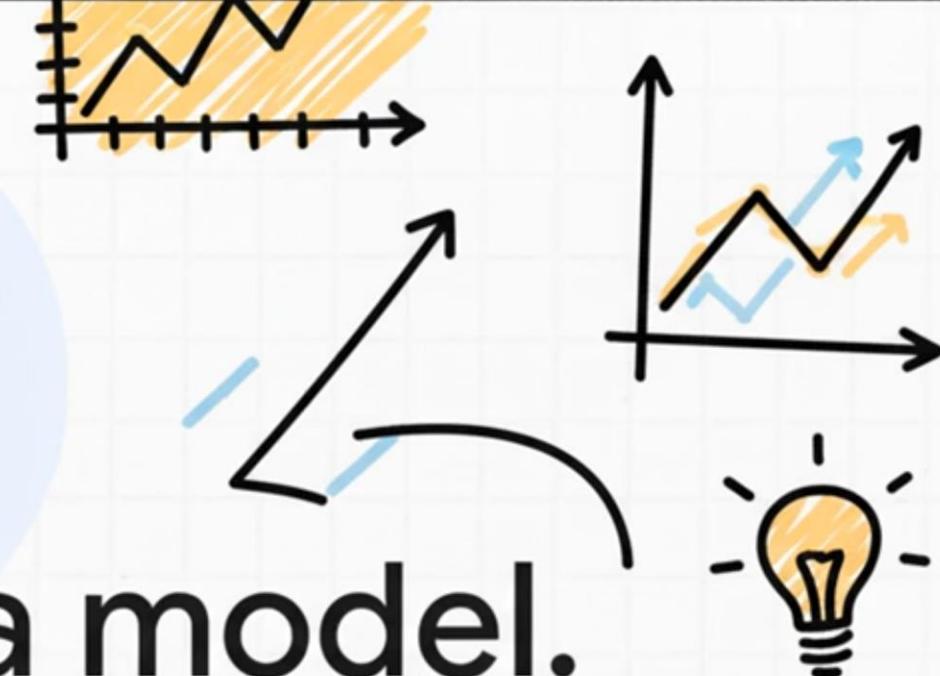
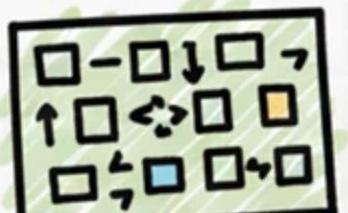
# • Regression Classification



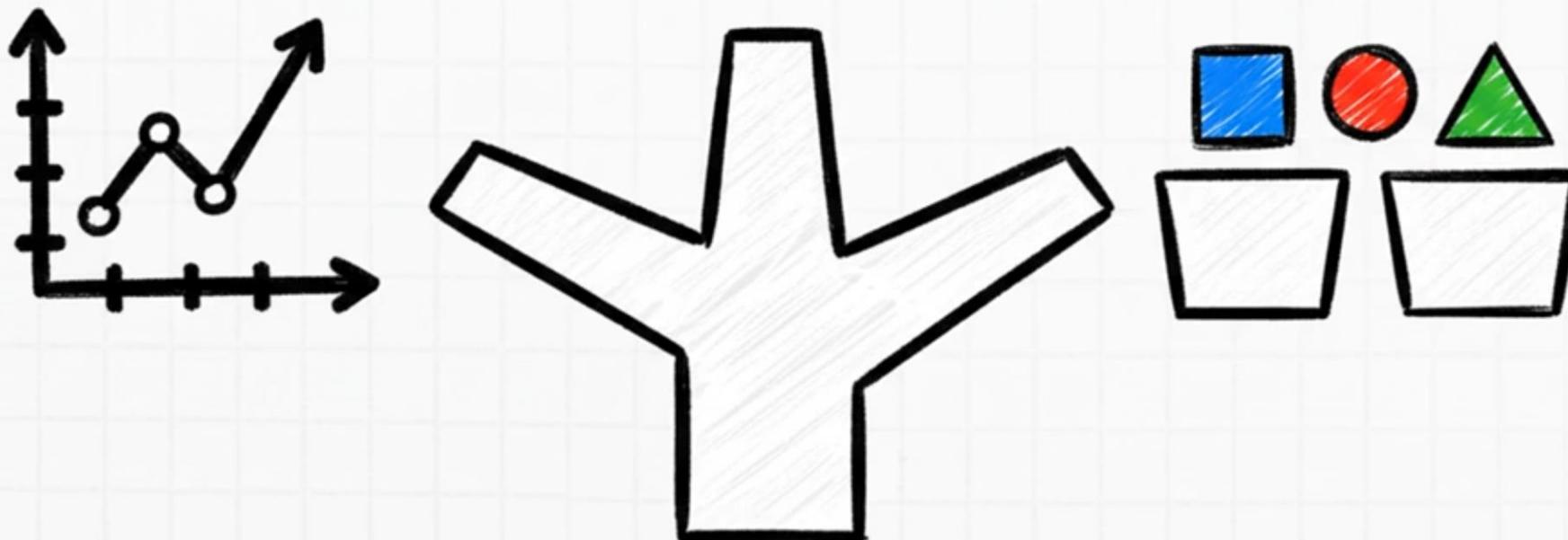


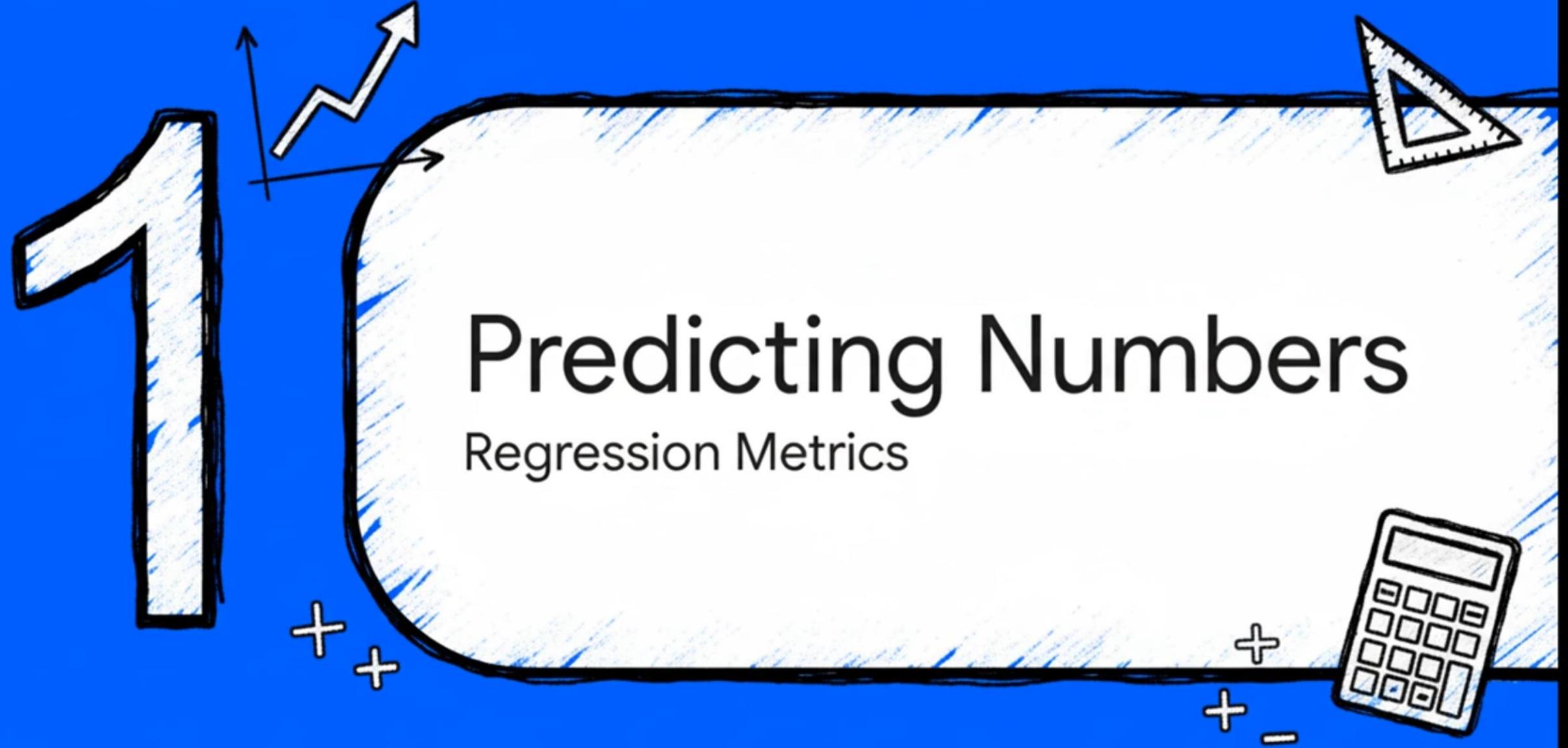
You've trained a model.

Now what?



Are you predicting a number  
or a category? The right way  
to measure your model  
depends entirely on its goal.





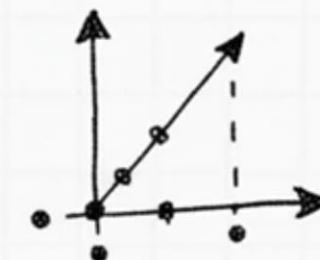
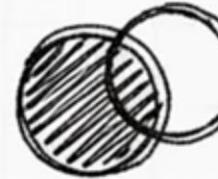
# Root Mean Squared Error (RMSE)

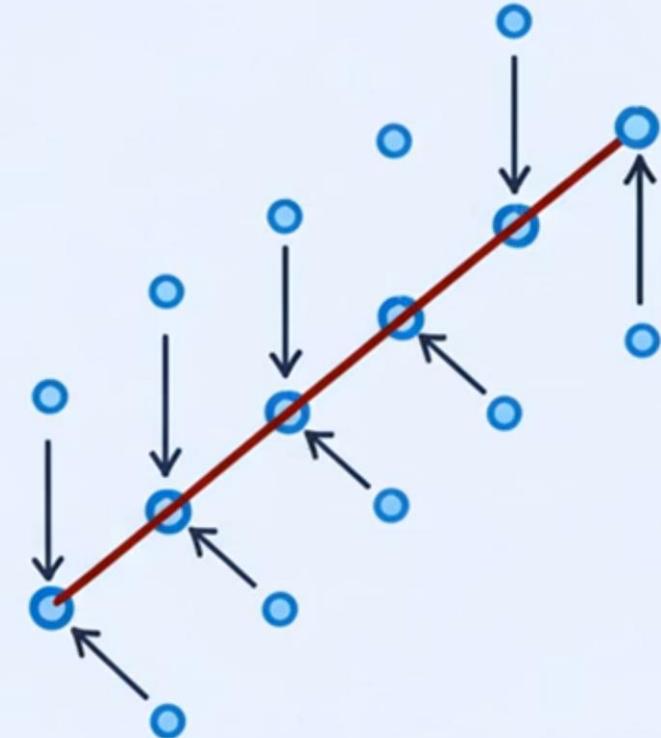
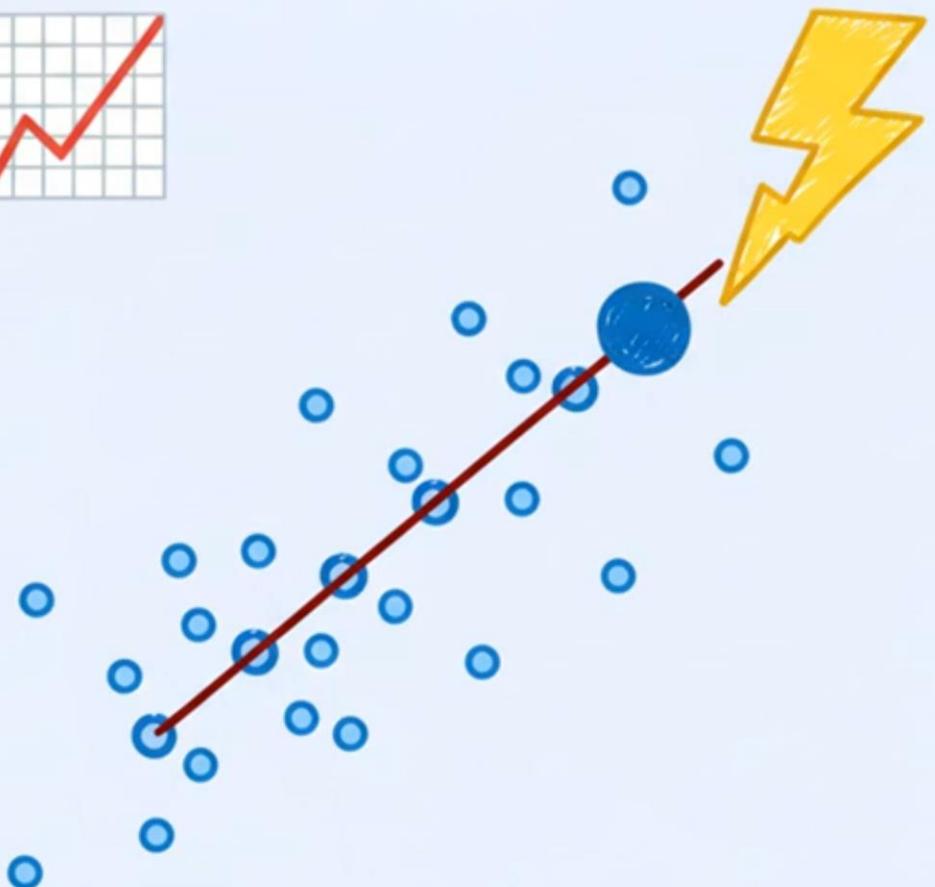
The typical size of your prediction error. It has the same units as your target and penalizes large errors.

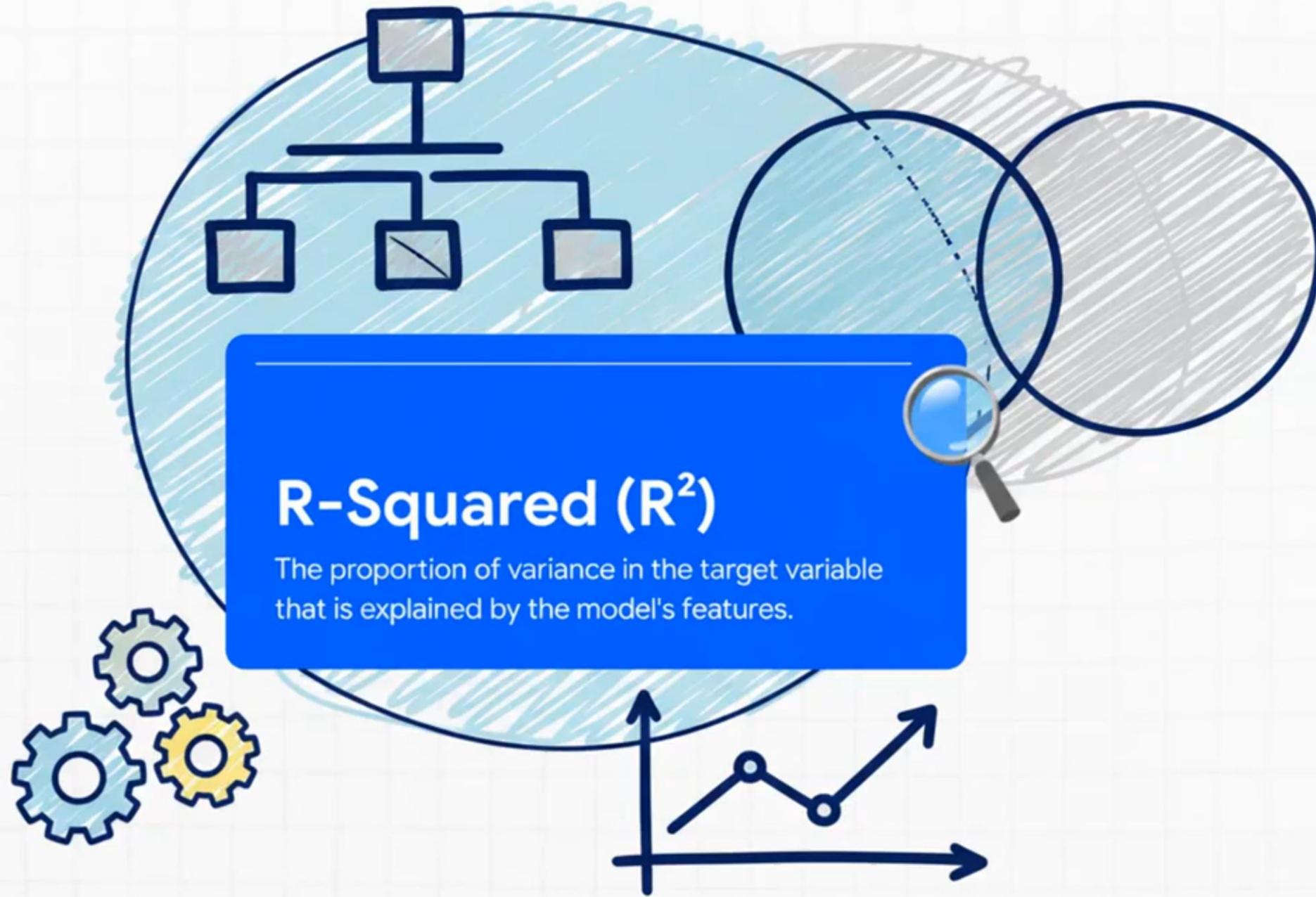


# Mean Absolute Error (MAE)

The average size of your prediction error. It's straightforward and less sensitive to outliers than RMSE.



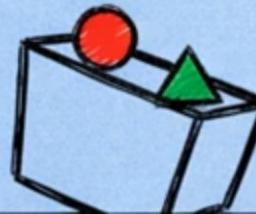
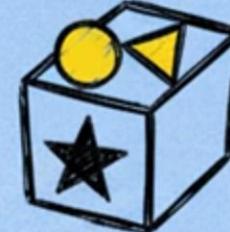
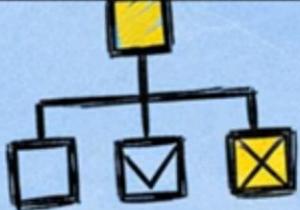
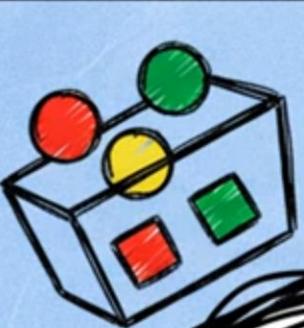
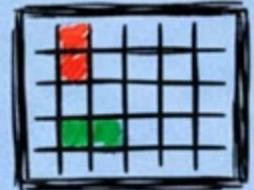




# 2

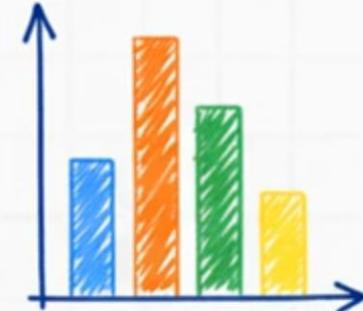
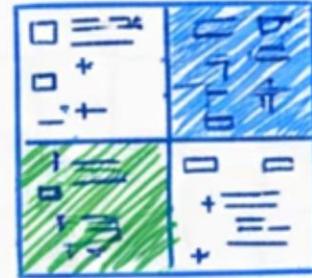
# Predicting Categories

Classification Metrics



# Accuracy

The simplest metric: the proportion of correct predictions out of all total predictions.



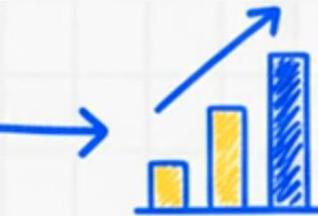
NotebookLM



DECODED  
DATA SCIENCE



“A model that predicts  
'not spam' for every email  
is **99% accurate**, but not  
actually **useful**.



NotebookLM

# Precision

When the model predicts 'yes,' how often is it correct? High precision means few false alarms.



AI NotebookLM



DECODED  
DATA SCIENCE

# Recall (Sensitivity)

Of all the actual 'yes' cases, how many did the model catch? High recall means few misses.





# 3

## In Practice

### Calculating Metrics



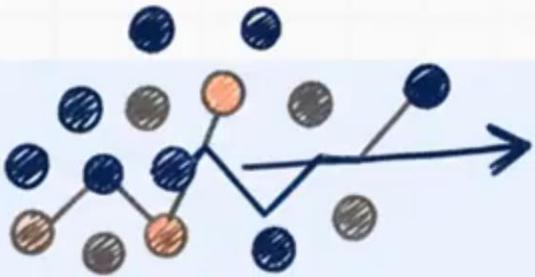
real-world goals

□	x	□
□	x	□
□	□	□
□	□	□

Classification



# How to Calculate



## Step 1: Predict

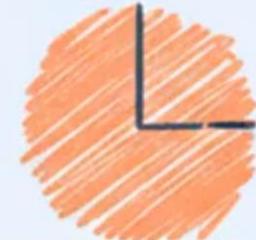
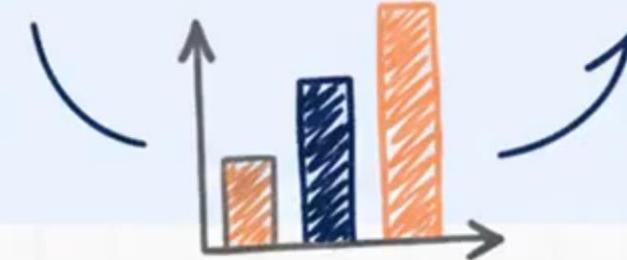
Get predictions from your model on the test data.

## Step 2: Import

Import metric functions from a library like 'sklearn.metrics'.

## Step 3: Calculate

Compare the predictions to the true values to get your score.



Libraries like **scikit-learn** make it easy. Just import `r2_score` or `precision_score` and apply them to your **data**.



0.92?

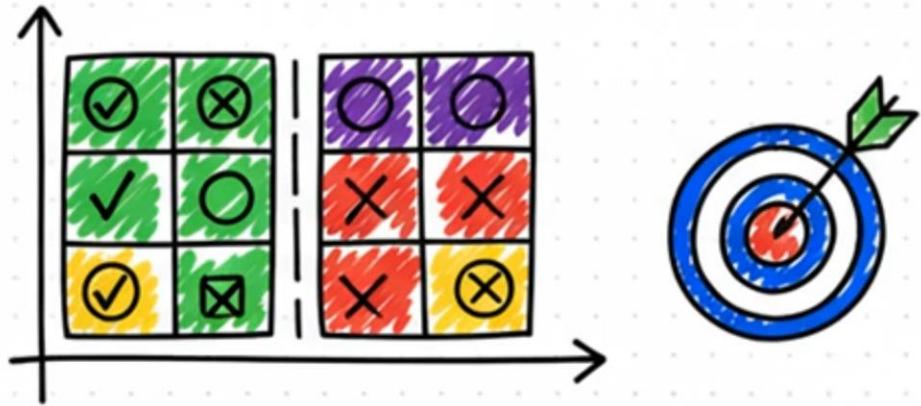
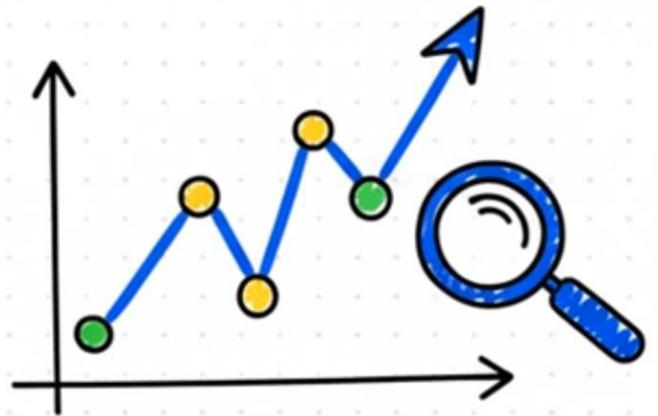


NUMBER WITHOUT CONTEXT

APPLYING CONTEXT



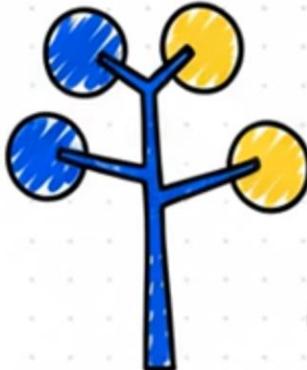
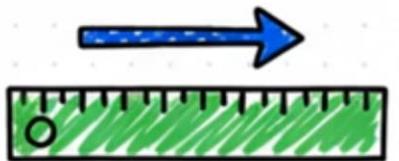
POWER PLANT  
EXAMPLE

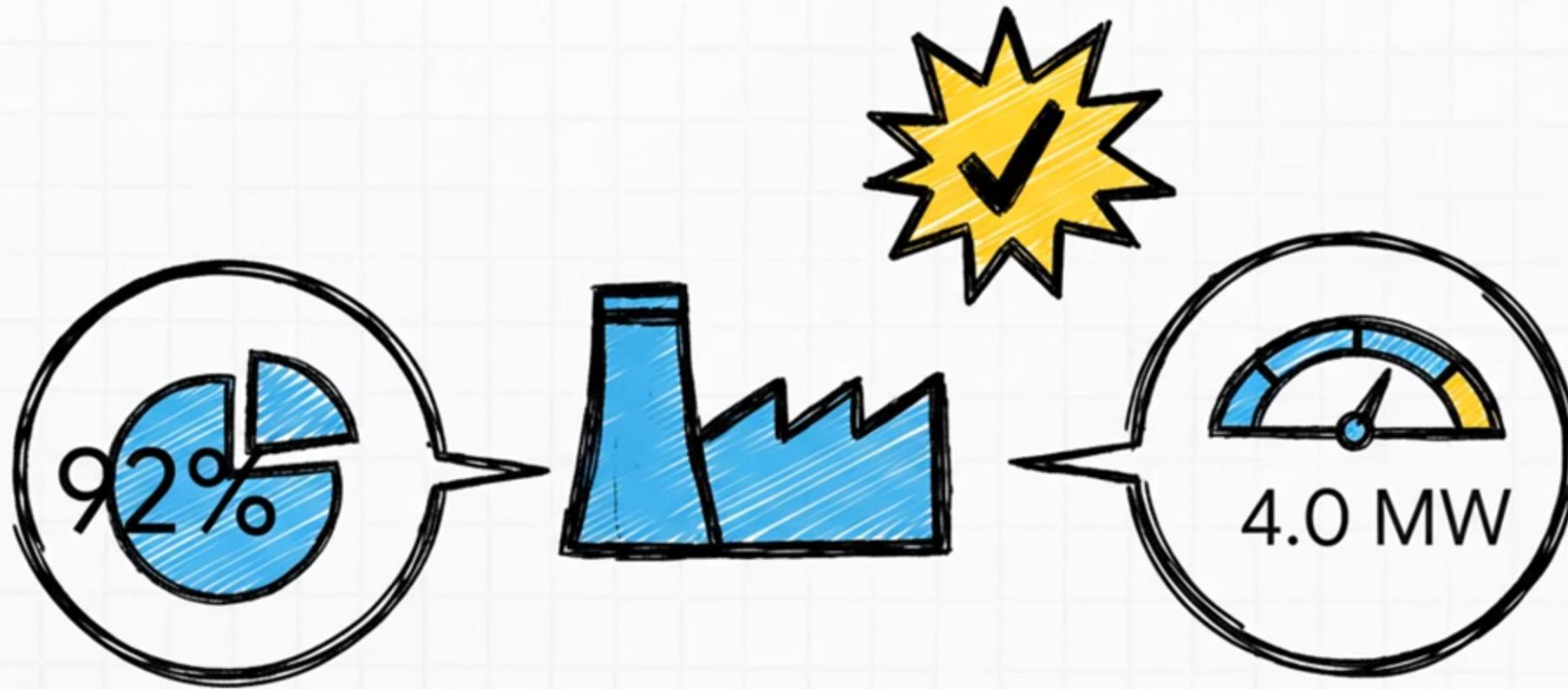


4

9

MW

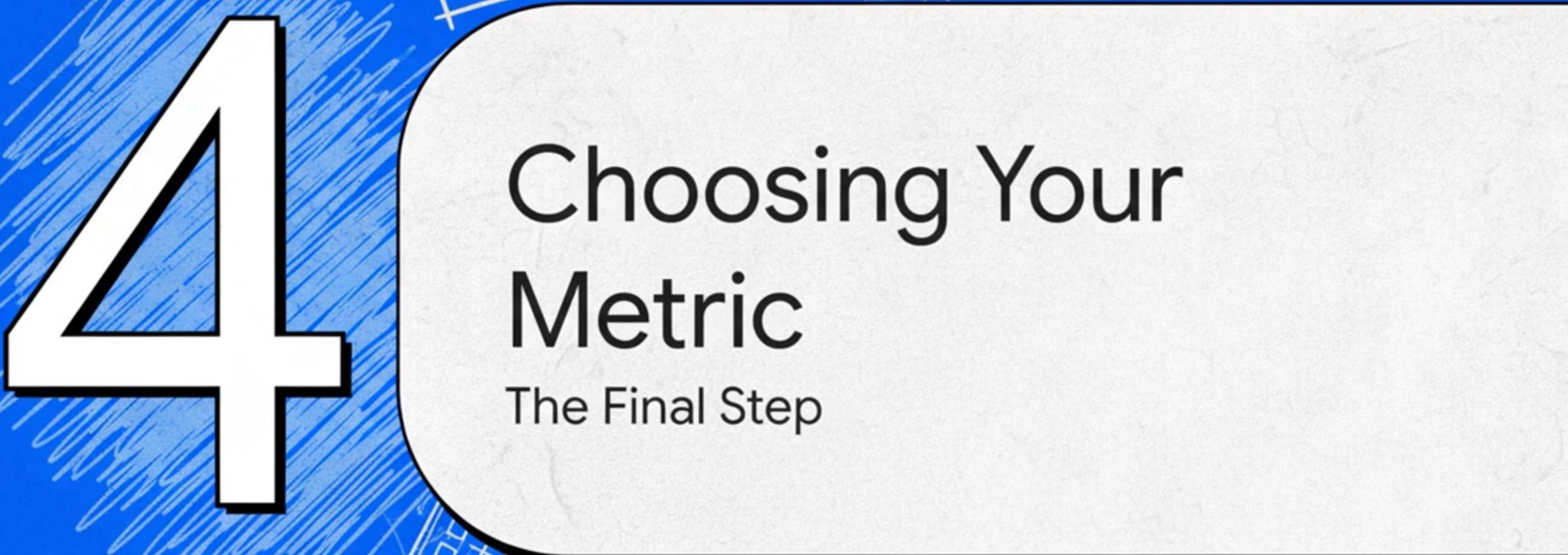


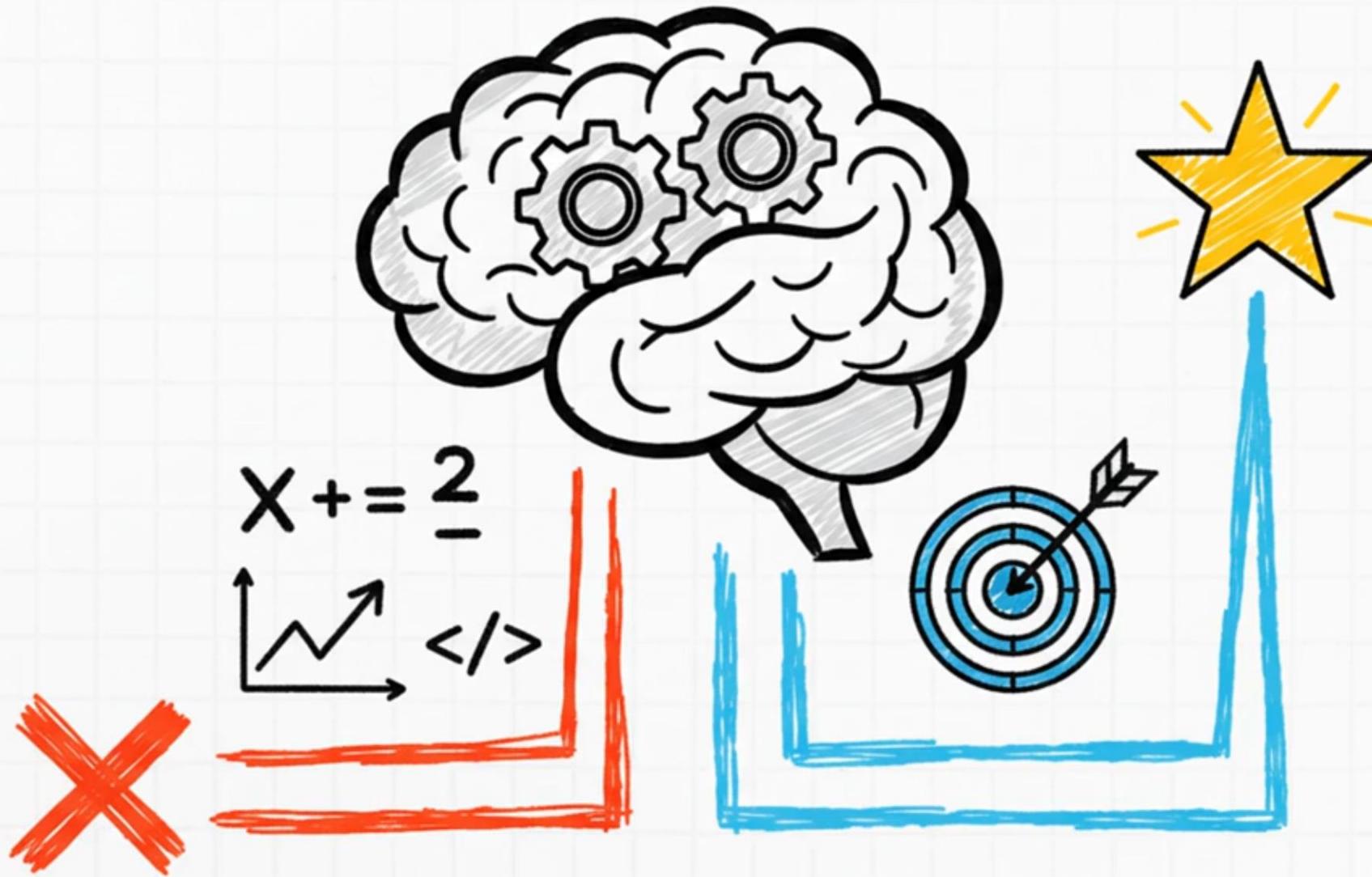


# 4

## Choosing Your Metric

The Final Step







For a medical  
diagnosis, what's  
worse: a **false positive**  
or a **false negative?**



