1. Leave-One-Out Cross-Validation (LOOCV):

In LOOCV, the dataset is divided into n subsets, where n is the number of samples in the dataset. For each iteration, one data point is retained as the validation set while the rest are used to train the model. This process is repeated n times.

Example Scenario: Suppose you have a dataset of 100 images of handwritten digits (0-9) and you want to train a classifier to recognize these digits.

Implementation:

- For each image in the dataset:
    - Train the classifier using all images except the current one.
    - Test the classifier on the current image.
- Calculate the overall accuracy of the classifier based on its performance on all images - by averaging out the score.

2. K-Fold Cross-Validation:

In K-Fold Cross-Validation, the dataset is divided into k subsets (folds) of approximately equal size. The model is trained k times, each time using $k-1$ folds for training and the remaining fold for validation.

Example Scenario: You have a dataset of 200 emails labelled as spam or non-spam, and you want to build a classifier to classify emails.

Implementation:

- Divide the dataset into k=5 folds (40 emails per fold). For each fold:
    - Train the classifier using the remaining 4 folds.
    - Test the classifier on the current fold.
- Calculate the overall accuracy of the classifier based on its performance on all folds - by averaging out score for each fold.

3. Validation Set Approach:

In the Validation Set Approach, the dataset is divided into two parts: a training set and a validation set. The model is trained on the training set and evaluated on the validation set. This process helps in tuning hyperparameters and assessing model performance.

Example Scenario: You want to train a regression model to predict housing prices based on various features.

Implementation:

Split the dataset into a training set (e.g., 80% of the data) and a validation set (e.g., 20% of the data).
Train the regression model on the training set using different hyperparameters.
Evaluate the model's performance on the validation set.
Choose the hyperparameters that result in the best performance on the validation set.
Optionally, retrain the model using the chosen hyperparameters on the entire dataset for final deployment.

Bootstrapping

https://www.youtube.com/watch?v=Xz0x-8-cgaQ&ab_channel=StatQuestwithJoshStarmer

Real-Life Example: Estimating the Mean Height of Students in a School
Step-by-Step Process:

Data Collection:

Collect a sample of heights from a subset of students in the school. Let's say we have measured the heights of 100 students.
Initial Calculation:

Calculate the mean height of the sample. Let's say the mean height of the sample is 165 cm.
Bootstrapping:

Define the number of bootstrap samples you want to create. Let's say we choose to create 1,000 bootstrap samples.

For each bootstrap sample:
Randomly select 100 heights from the original sample with replacement. This means some heights may be selected multiple times, and some may not be selected at all.
Calculate the mean height for this bootstrap sample.
After creating 1,000 bootstrap samples, you will have 1,000 bootstrap means.

Statistical Analysis:
Analyze the distribution of the 1,000 bootstrap means. You can plot a histogram to visualize the distribution.
Calculate statistics such as the mean, standard deviation, and confidence intervals of the bootstrap means.

Inference:

Use the distribution of bootstrap means to make inferences about the population mean height. For example, you can calculate the 95% confidence interval based on the bootstrap distribution. This interval gives you a range within which you are 95% confident that the true population mean height lies.
Interpretation:

Interpret the results in the context of the original question. For instance, you might conclude that based on the sample data, the mean height of students in the school is estimated to be 165 cm, with a 95% confidence interval of [164 cm, 166 cm].