Cross validation:

Step-1: You will split the data into training and testing – 2 data sets in simple

Step-2: You build the model on training and test the model performance in test data (which is unseen data)

Step-3: You will verify the model accuracy in both training and testing.

* First you try to improve the model performance in train set by changing the model parameters.
* When you have achieved a decent accuracy in training, you will move the model to the test data for predictions.
* Now check the model accuracy in test data. If the model accuracy doesn’t match the training accuracy, then change the model (build a new model in train set by changing parameters…until train and test accuracies match ~ approximately same)
* You can also perform K-fold validation or split the data into 60-20-20 (train, test, validation datasets)
* While testing the model in test dataset you create a confusion matrix and check metrics like **Accuracy, Precision, Recall, F1 Score and Support.**

|  |  |  |
| --- | --- | --- |
|  | **Predicted** |  |
|  |  | **Positive** | **Negative** |
| **Observed** | **Positive** | TP (# of TPs) | FN (# of FNs) |
|  | **Negative** | FP (# of FPs) | TN (# of TNs) |

1. **TN / True Negative:**  case was negative and predicted negative
2. **TP / True Positive:**  case was positive and predicted positive
3. **FN / False Negative:**  case was positive but predicted negative
4. **FP / False Positive:**  case was negative but predicted positive

**Accuracy:**

Accuracy (ACC) is calculated as the number of all correct predictions divided by the total number of the dataset.

\mathrm{ACC = \displaystyle \frac{TP +TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}}

**2. What percent of positive predictions were correct?**

How many of the positively classified were relevant. A test can cheat and maximize this by only returning positive on one result it’s most confident in.

**Precision (Positive predictive value):**

Precision (PREC) is calculated as the number of correct positive predictions divided by the total number of positive predictions

The best precision is 1.0, whereas the worst is 0.0

**3. What percent of the positive cases did you catch?**

How good a test is at detecting the positives. A test can cheat and maximize this by always returning “positive”.

**Recall (Sensitivity or True positive rate):**

Recall (REC) is calculated as the number of correct positive predictions divided by the total number of positives. It is also called Sensitivity (SN) or true positive rate (TPR).

recall = tp / t = tp / (tp + fn) The best sensitivity is 1.0, whereas the worst is 0.0

1. **Specificity –** how good a test is at avoiding false alarms? A test can cheat and maximize this by always returning “negative”.

**F-score is a harmonic mean of precision and recall.**\mathrm{F_{\beta} = \displaystyle \frac{(1 + \beta^2) (PREC \cdot REC)}{(\beta^2 \cdot PREC + REC)}} **(Less used)**