ML Evaluation Measures

20CP401T

Himanshu K. Gajera
Department of Computer Science & Engineering
Pandit Deendayal Energy University, Gandhinagar

Evaluation Measures

Success: actual output = target

Error: actual output ≠ target

%error = #errors / #samples

Evaluation Measures

Training: In training the model is built
Testing: In testing, the model is applied to the new data.

The goal in building a machine learning algorithm is to perform well on both training and testing data.

Error on training data is called training error.
Error on test data is called test error.
Test error indicates how well model will perform on new data.(**Generalization**)

Performs well on new data ⇒ Good generalization
Test error = Generalization error

Overfitting and Underfitting

- **Overfitting**: If a model has low training error and high test / generalization error, then it is called overfitting

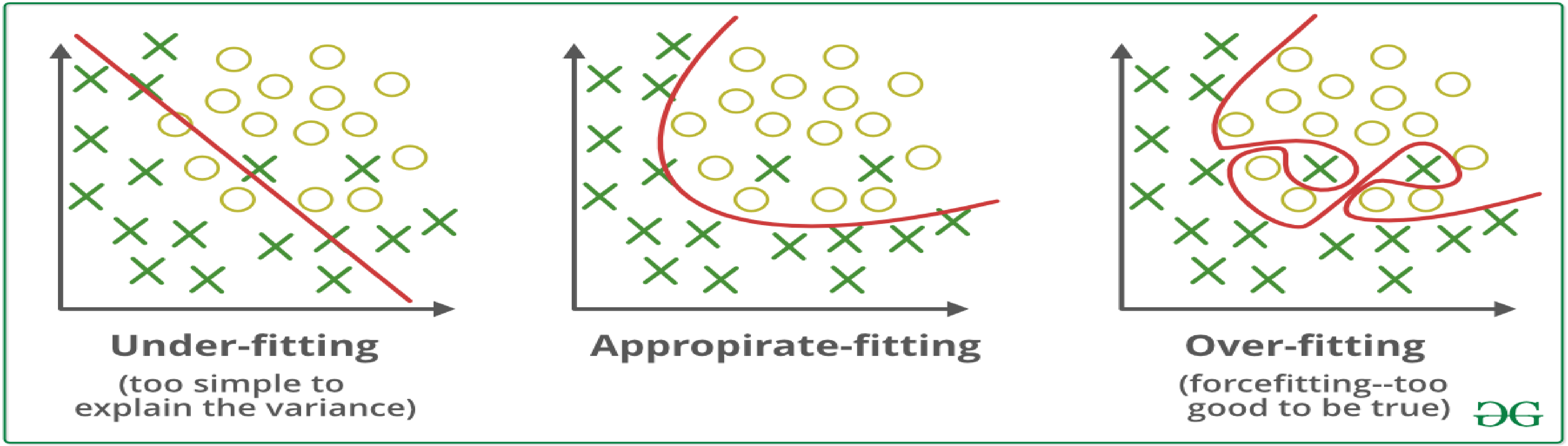
- Fit to the noise in the training data

    - **Overfitting = Poor generalization**

- **Underfitting**: high training error and high test / generalization error

- The cause of the poor performance of a model in machine learning is either overfitting or under fitting the data.

Overfitting and Underfitting

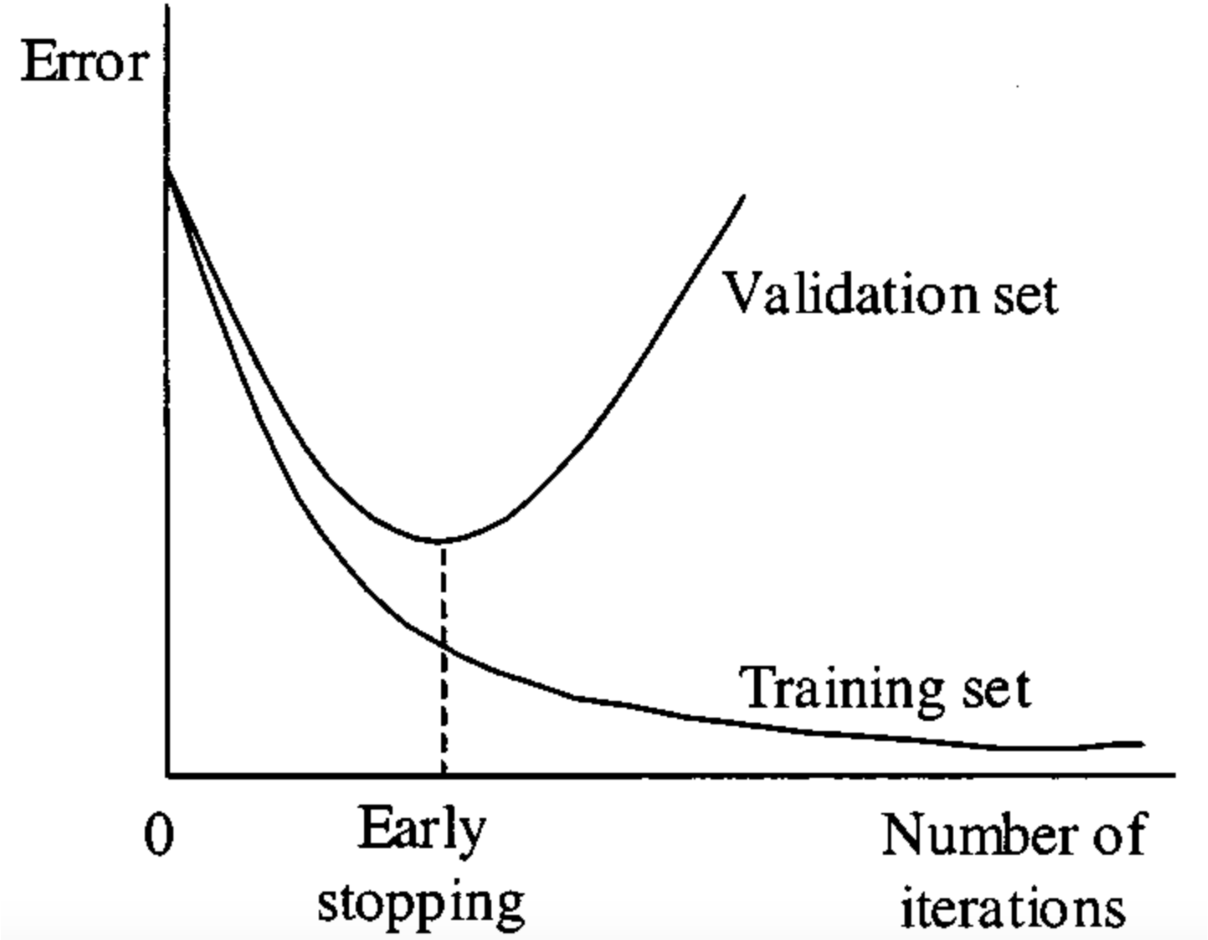


What causes overfitting?

It occurs when a model is too complex i.e., it has too many parameters related to the number of training samples.

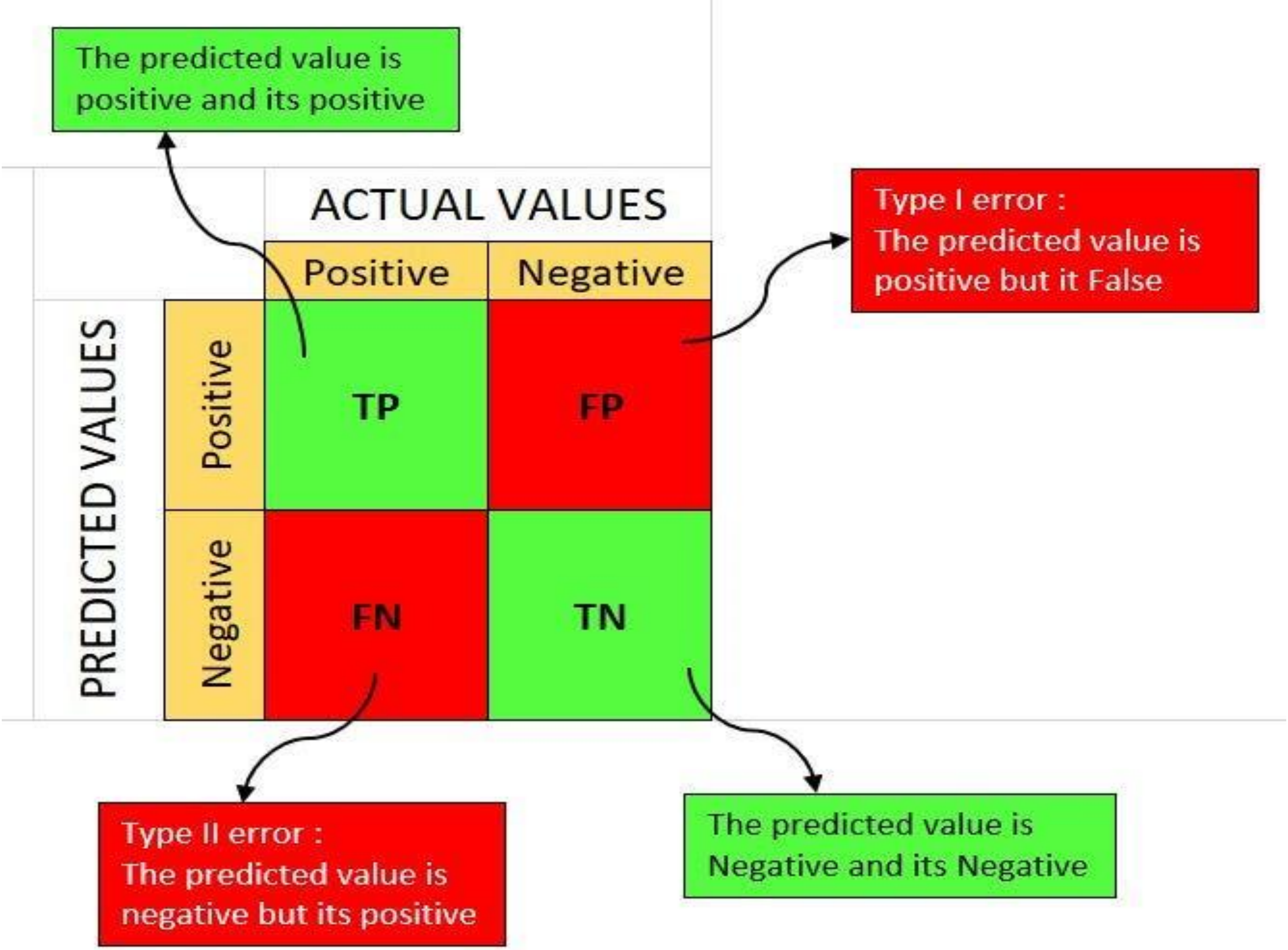
So, to avoid overfitting the model need to be kept as simple as possible.

How to prevent Overfitting?



Confusion matrix

- Confusion Matrix is a tool to determine the performance of classifier. It contains information about actual and predicted classifications.



Confusion matrix

- True Positive(TP): # positive samples correctly identified as positive.
- False Negative(FN): #positive samples incorrectly identified as negative
- False Positive(FP): # negative samples incorrectly identified as positive
- True Negative(TN): # negative samples correctly identified as negative

Accuracy, sensitivity, specificity

Sensitivity = True positive rate
= recall
= 1 – False negative rate
= TP / (TP + FN)

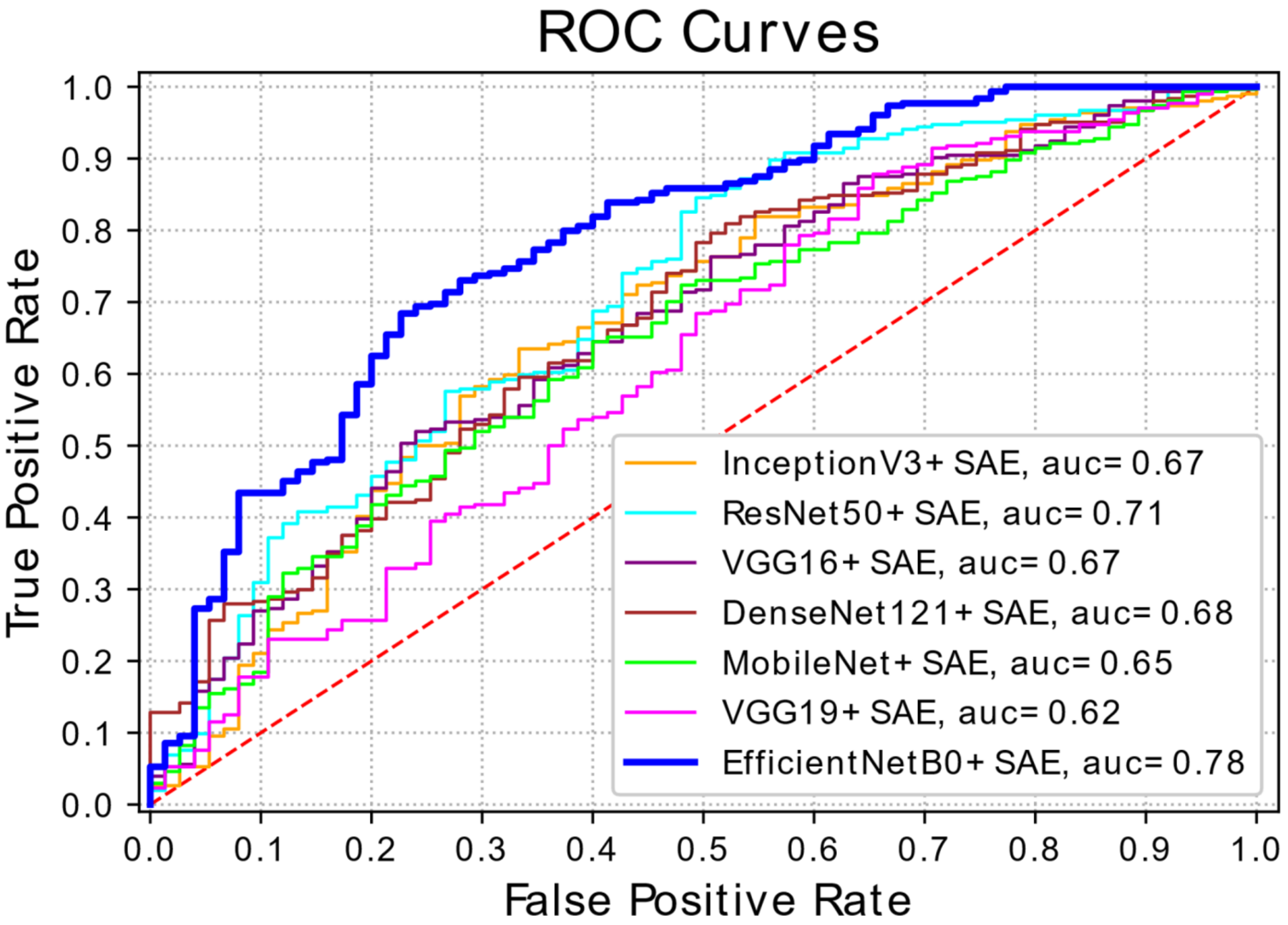
Specificity = True negative rate
= 1 – False positive rate
= TN / (TN + FP)

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{recall}}{\text{Precision} + \text{recall}}$$

ROC Curve (Receiver Operating Characteristic Curve):



Confusion matrix

Predicted Class

Actual Class		Positive	Negative	
	Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

Regression Error Metrics

Mean Absolute Error: average of the difference between the Original Values and the Predicted Values. It gives us the measure of how far the predictions were from the actual output.

Divide by the total number of data points

Predicted output value

Actual output value

$$MAE = \boxed{\frac{1}{n}} \sum \underbrace{\left| \boxed{y} - \boxed{\hat{y}} \right|}_{\text{The absolute value of the residual}}$$

Sum of

$$\mathrm{MAE} = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$

Regression Error Metrics

Mean Square Error(MSE): average of the square of the difference between the original values and the predicted values. The advantage of MSE being that it is easier to compute the gradient, where as Mean Absolute Error requires Complicated linear programming tools to compute the gradient

$$\mathrm{MSE} = \overbrace{\boxed{\frac{1}{n}\sum_{i=1}^{n}}}^{\text{Mean}} \Big(\overbrace{\boxed{Y_i - \hat{Y}_i}}^{\text{Error}}\Big)^{\overbrace{\boxed{2}}^{\text{Squared}}}$$

Root Mean Square Error (RMSE):

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

Regression Error Metrics

Mean absolute percentage error (MAPE):

$$MAPE = \frac{\overset{\text{Multiplying by 100\% converts to percentage}}{\boxed{100\%}}}{n} \sum \left| \frac{\overbrace{y - \hat{y}}^{\text{The residual}}}{\underset{\text{Each residual is scaled against the actual value}}{\boxed{y}}} \right|$$

Regression Error Metrics

Mean percentage error :

$$MPE = \frac{100\%}{n} \Sigma \left(\frac{y - \hat{y}}{y} \right)$$

Sum of Squared Errors

It is a common metric used to quantify the difference between the predicted values and the actual values in a regression problem. Sum of Squared Errors is often used in training and evaluating regression models to measure how well the model fits the training data.

$$\underbrace{\sum_{i=1}^{\overset{\text{number of samples}}{\boxed{n}}} \left(\overset{\text{real value}}{\boxed{Y_i}} - \overset{\text{predicted value}}{\boxed{\hat{Y}_i}}\right)^2}_{\text{sum of the errors of all samples}}$$

The goal in regression is to minimize the SSE.