MSE491: Application of Machine Learning in Mechatronic Systems

Skewed Classes &
Debugging A ML Algorithm

Mohammad Narimani, *Ph.D.*, *P.Eng.*
Lecturer
School of Mechatronic Systems Engineering
Simon Fraser University

What is an accurate model in Machine Learning?

- The most challenging question in developing a machine learning model is that “how to get the most accurate model?”

Taken from www.analyticsvidhya.com

What is Skewed Classes?

- **Skewed classes** or **Skewed dataset** is referring to a dataset when difference between the numbers of examples belonging to each class is large. This leads to some difficulties for a learning system to learn the concept related to the minority class.

- Let's learn Skewed classes with a simple Example:
  - Statistical information have shown that less than 0.5% of the population get a specific cancer. That means if we close our eyes and say nobody got that type of cancer, it has 99.5% accuracy.
  - Now, consider that a classification algorithm has been developed for cancer detection. The obtained results have shown that the trained model accuracy is 99.1%!!

$$accuracy = \frac{TP+TN}{TP+FN+TN+FP}$$

Conclusion: Without a diagnostic tool (just closing our eyes) the prediction accuracy is higher!!!

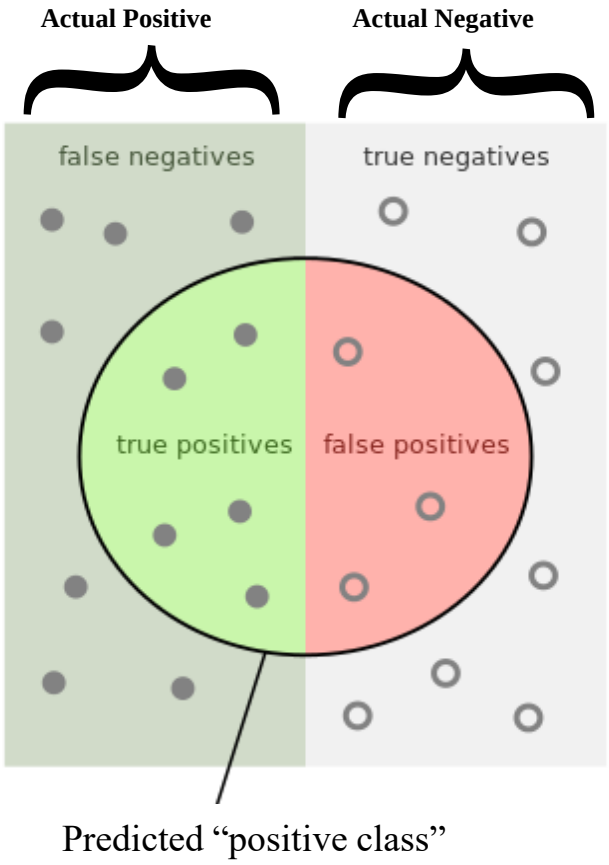
Taken from www.analyticsvidhya.com

Confusion Matrix

- Consider a binary dataset with "positive" and "negative" classes.

- Assume that using a classifier (a ML algorithm) a class is predicted, say "positive class".

	Actual	
Predicted		



Taken from www. wikipedia.org

Confusion Matrix

- $accuracy = \frac{TP+TN}{TP+FN+TN+FP}$
- $Precision = \frac{TP}{TP+FP} =$  $TP + FP$ *is all predicted positive*
- $Recall(sensitivity) = \frac{TP}{TP+FN} =$  $TP + FN$ *is all actual positive*
- $Specifity = \frac{TN}{TP+FP}$

Taken from www.analyticsvidhya.com

F1 score

- F1 score is a function of *Precision* and *Recall*

- The F1 score tries to give more weight to false negatives and false positives when there is an uneven class distribution (large number of actual negative)

- It makes a balance between Precision and Recall

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Debugging A ML Algorithm

- Suppose that a regularized algorithm is trained, and the resulting model has remarkable error in detection (prediction). We need to modify the model by following steps:

Debugging A ML Algorithm

- Choosing the degree of polynomial in a model:

Debugging A ML Algorithm

- Choosing the rergularization parameter in the cost function:

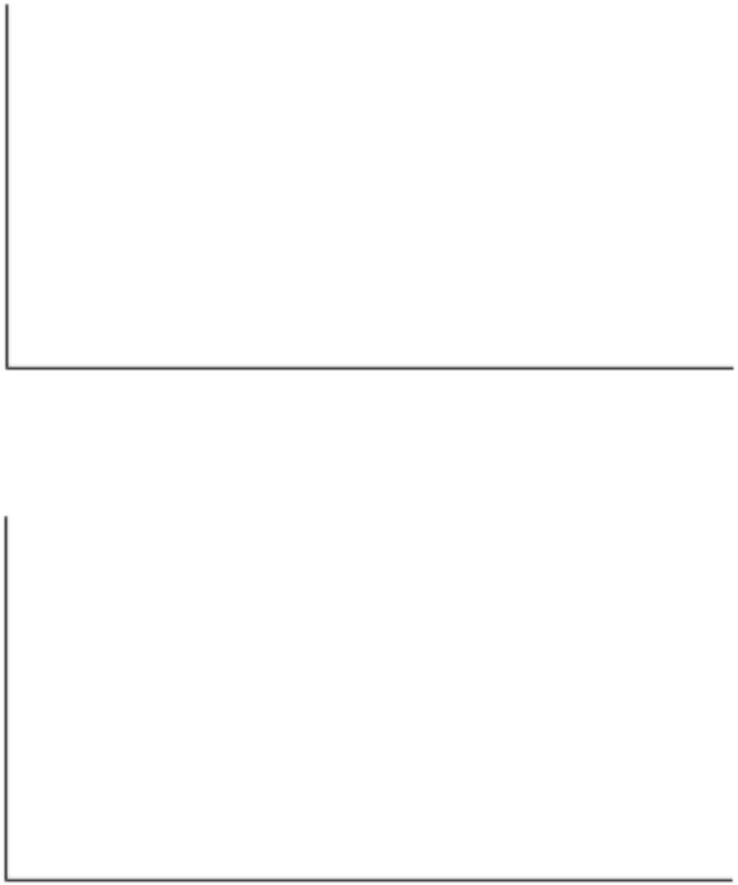
Learning Curves

- Learning curves are plots of a cost function changes (error changes) as the training set size increases.

- It is diagnostic tool in machine learning for the evaluation of a trained model.

Learning Curves

- Learning curves show how error for training dataset and validation dataset change as the training set size increases.



Learning Curves

- It is diagnostic tool in machine learning for algorithms that learn from a training dataset.

- A model is evaluated by calculating the error on the training dataset and a validation dataset after each update (increase the number of training samples) during training.