Model Selection

Presented by Blessing Itoro Bassey

African Master's in Machine Intelligence

December 4, 2019

Outlines

1. INTRODUCTION
   - What Model is
   - The Trio
2. Model Evaluation
   - Classification Related Metrics
   - Regression Related Metrics
   - Performance Metric
3. Model Selection
   - Data Spliting and hyperparameter tuning
   - Algorithm Selection
   - Hyperparameter Selection
4. Summary and Conclusion

What Model is

To model is to create a representation or copy of something, often on a smaller scale, or to base your method of behavior on another person or procedure.

Model, Algorithm and Hypothesis

- **Machine Learning model**: mathematical representation of a real-world process.

Model, Algorithm and Hypothesis

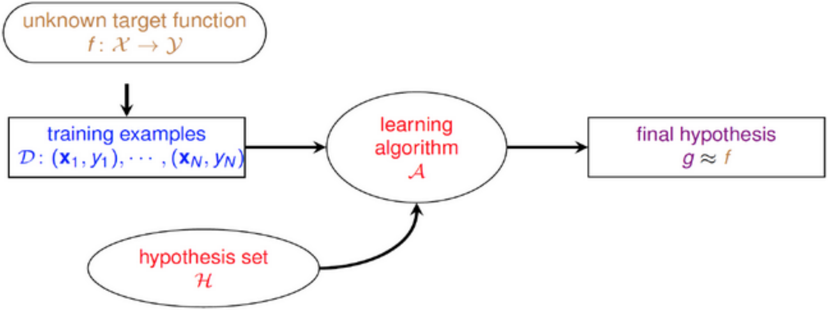
- **Machine Learning model**: mathematical representation of a real-world process.
- **Machine Learning Algorithm:** programs that can learn from data and improve from experience, without human intervention.

Model, Algorithm and Hypothesis

- **Machine Learning model**: mathematical representation of a real-world process.
- **Machine Learning Algorithm:** programs that can learn from data and improve from experience, without human intervention.
- **Machine Learning Hypothesis:** class of candidate model that approximates a target function for mapping examples of inputs to outputs.

Model, Algorithm and Hypothesis

It can be shown in the following diagram:



Model Selection

When approaching some practical problem, we can think of several algorithms that may yield a good solution, each of which might have several parameters. Hence we have the following questions to battle with:

- How can we choose the best algorithm for the particular problem at hand?
- How do we set the algorithm's parameters?.

Model Selection

When approaching some practical problem, we can think of several algorithms that may yield a good solution, each of which might have several parameters. Hence we have the following questions to battle with:

- How can we choose the best algorithm for the particular problem at hand?
- How do we set the algorithm's parameters?.

The task of answering the above questions is known as model selection. In general, **model selection** is the task of selecting an appropriate model for the learning task based on the data itself.

Model Evaluation

"Why do we care about performance estimates at all?"

Ideally, the estimated performance of a model tells how well it performs on unseen data. Since we are typically interested in selecting the best-performing model from the set of models that we have, we need to find a way to estimate their respective performances in order to rank them against each other.

Classification Related Metrics

Classification is one of the most widely used problems in machine learning with various industrial applications,

Models such as support vector machine (SVM), logistic regression, decision trees, random forest, XGboost, convolutional neural network[1], recurrent neural network are some of the most popular classification models.

Classification Accuracy

Let's assume we are building a binary classification to classify cat images from non-cat images. And let's assume our test set has 1100 images (1000 non-cat images, and 100 cat images), with the below confusion matrix.

		Actual Class	
		Cat	Non-Cat
Predicted Class	Cat	90	60
	Non-Cat	10	940

Figure: A sample confusion matrix

Classification accuracy$= \dfrac{90+940}{1000+100} = 93.6\%$

Precision

Precision= True_Positive/ (True_Positive+ False_Positive)

- **Precision_cat** $= \frac{90}{90+60} = 60\%$
- **Precision_NonCat** $= \frac{940}{950} = 98.9\%$

Accuracy refers to how close a measurement is to the true value. Precision is how consistent results are when measurements are repeated.

Sensitivity and Specificity

Sensitivity and specificity are two other popular metrics mostly used in medical and biology related fields, and are defined as:

- **Sensitivity/Precision** $= \dfrac{TP}{TP + FN}$

- **Specificity/Recall** $= \dfrac{TN}{TN + FP}$

It's important to recognize that **sensitivity** and **specificity** exist in a state of balance. Increased sensitivity – the ability to correctly identify people who have the disease — usually comes at the expense of reduced specificity (meaning more false-positives).

ROC Curve

The receiver operating characteristic curve is plot which shows the performance of a **binary classifier** as function of its cut-off threshold. It essentially shows the **true positive rate** (TPR) against the **false positive rate** (FPR) for various threshold values.

As an example your model may predict the below probabilities for 4 sample images: [0.45, 0.6, 0.7, 0.3]. Then depending on the threshold values below, you will get different labels:

cut-off= 0.5: predicted-labels= [0,1,1,0]
cut-off= 0.2: predicted-labels= [1,1,1,1]
cut-off= 0.8: predicted-labels= [0,0,0,0]

ROC

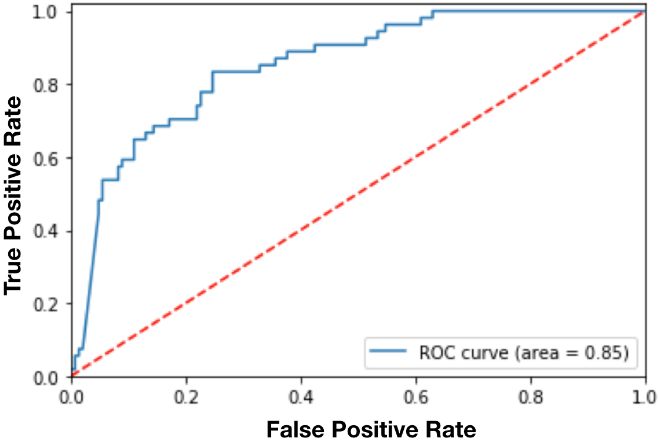


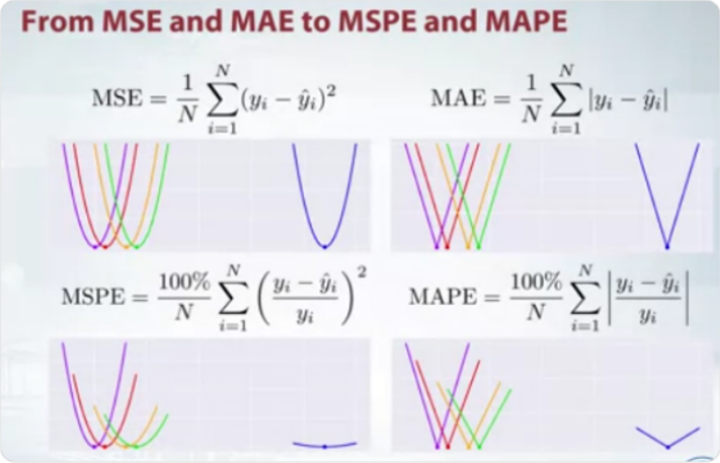
Figure: ROC Curve

Regression Related Metrics

Regression models are another family of machine learning and statistical models, which are used to predict a continuous target values. With a wide range of applications, from house price prediction, E-commerce pricing systems. etc.

Models such as linear regression, random forest, XGboost, convolutional neural network, recurrent neural network are some of the most popular regression models.

Performance Metric

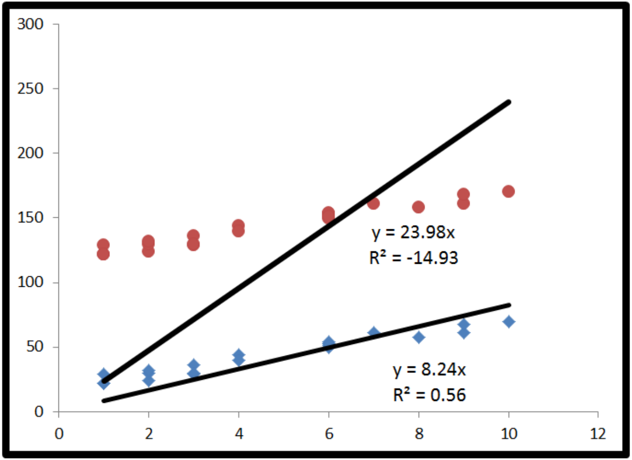


R Squared Formula

R-Squared: It looks at how closely the data fits to the regression line, with a high r-squared meaning the model better fits the data. r-squared falls between 0 (low) and 1 (high).

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}}$$

R Squared Formula



The cons of R Squared Formula

This measure tends to favour more complex models, it doesn't actually penalise you for overfitting. You can crank the model up as far as you want. The more features you add, the better your R Square.

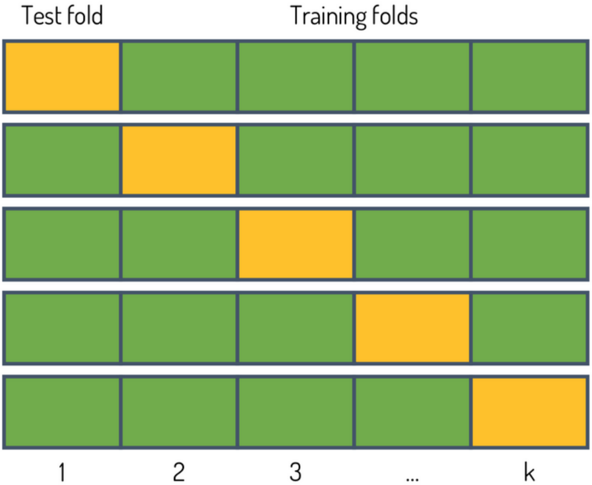
Model Selection and Hyperparameter Tuning

The recommended strategy for model selection depends on the amount of data available. If ***plenty of data*** is available, we may split the data into several parts, each serving a special purpose. For instance, for ***hyperparameter tuning*** we may split the data into three sets: ***train / validation / test.***

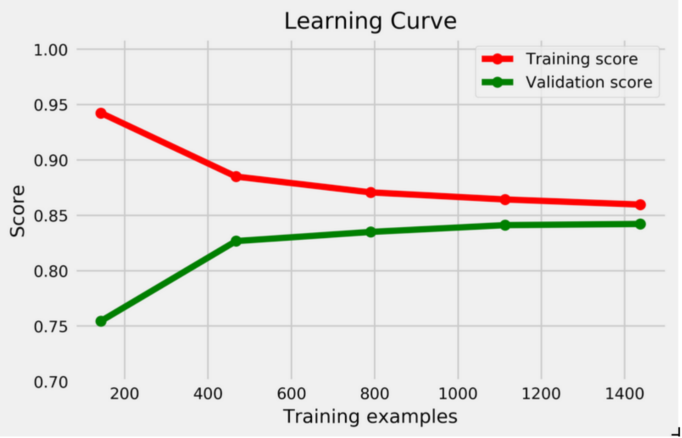
Hyperparameter Tuning

At the tuning stage, selecting a model based on the validation set performance seems to be a reasonable approach. For hyperparameter selection, we can use K-fold cross-validation (CV). Here, the main idea behind cross-validation is that each sample in our dataset has the opportunity of being tested.

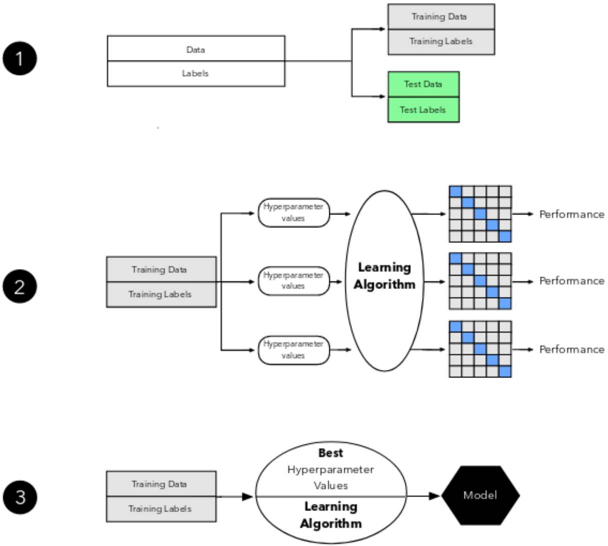
Illustration of 5 fold cross validation



Learning Curves



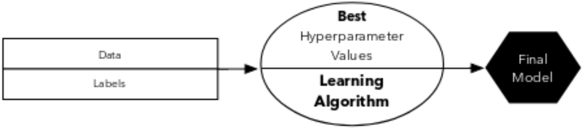
Summary and Conclusion



Summary and Conclusion

4

5



Reference

- https://sebastianraschka.com/pdf/manuscripts/model-eval.pdf Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. Sebastian Raschka University of Wisconsin–Madison Department of Statistics November 2018
- https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3e282a2ce

Reference

- http://www.fairlynerdy.com/what-is-r-squared/
What Is R Squared And Negative R Squared
- https://www.programcreek.com/2013/01/what-is-machine-learning/
What is Machine Learning?
- https://towardsdatascience.com/a-short-introduction-to-model-selection-bb1bb9c73376
- https://twitter.com/KirkDBorne/status/ 1154840570705170433

Thank You For your Kind Attention