Machine Learning
Evaluation of Learning Algorithms

Satishkumar L. Varma

Department of Information Technology
SVKM's Dwarkadas J. Sanghvi College of Engineering, Vile Parle, Mumbai.
ORCID | Scopus | Google Scholar | Google Site | Website

Outline

- Machine Learning Model
  - Evaluating a Learning Algorithm
  - Evaluating Hypothesis
  - Model Selection and Train/ Validation/ Test Sets
  - Bias Vs variance: Regularization and Bias/ Variance, Learning Curve, Error Analysis
  - Handling Skewed Data: Error Matrices for Skewed Classes
  - Trade-off between Precision and recall
- Model Evaluation
- Model Improvement (Ensemble Learning)
- Optimization

Analytical Models

- Models are used for prediction, description and prescription based on key elements in data.
- These **analytical models** are used for predicting, understanding and making data–driven decisions.
- **Predictive models**
  - Designed to make predictions or estimates about future events.
  - These ML models analyze historical data to identify patterns and estimates future outcome.
  - Predictive modelling is a process used in data science to create a mathematical model.
  - Such models predicts an outcome based on input data.
  - Examples: Regression analysis, decision trees, and neural networks.
- **Descriptive models**
  - Designed to provide the past and present, helping to identify patterns and trends in data.
  - Statistical analysis techniques to extract meaningful information from the data, such as mean, standard deviation, distribution, and correlations.
  - Examples: Clustering, association rule mining, and anomaly detection.
- **Prescriptive models**
  - Designed to provide recommendations or guidance based on the available data regarding future actions.
  - Examples: NLP

Predictive Modeling

- Dependent and independent variables are key concepts in predictive modeling and statistics.
- **Dependent Variable (DV)**
  - The dependent variable is the main factor or outcome that you're interested in predicting or understanding.
  - It's often denoted as "Y" in mathematical equations.
  - In a study or experiment, the DV is the variable that is measured or observed.
  - For example, in a study looking at the effect of studying time on test scores,
    - the test scores would be the DV because they depend on the amount of time spent studying.
- **Independent Variable (IV)**
  - IV are the factors or variables that are manipulated or controlled in a study.
  - They are used to predict or explain changes in the DV.
  - IV are often denoted as "X" in mathematical equations.
  - IV would be the amount of time spent studying,
    - as this is the variable that is being manipulated to see its effect on test scores.

Satishkumar L. Varma   www.sites.google.com/view/vsat2k

Predictive Modeling

- The two most commonly employed predictive modeling methods are **Regression and NN**.
- The accuracy of predictive analytics and every predictive model depends on
- Quality of your data
- Choice of variables

Comparison	Descriptive Models	Predictive Models
Basic	It determines, what happened in the past by analyzing stored data.	It determines, what can happen in the future with the help past data analysis.
Preciseness	It provides accurate data.	It produces results does not ensure accuracy.
Practical analysis methods	Standard reporting, query/drill down and ad–hoc reporting.	Predictive modelling, forecasting, simulation and alerts.
Require	It requires data aggregation and data mining	It requires statistics and forecasting methods
Type of approach	Reactive approach	Proactive approach
Describe	Describes the characteristics of the data in a target data set.	Carry out the induction over the current and past data so that predictions can be made.
Methods(in general)	what happened? where exactly is the problem? what is the frequency of the problem?	what will happen next? what is the outcome if these trends continue? what actions are required to be taken?

Model Accuracy

- **Model Accuracy**
  - There will always be a slight difference in what our model predicts and the actual predictions.
  - ML models allows machines to perform data analysis and make predictions.
  - However, ML models are not accurate and can make predictions errors.
- **Errors in Machine Learning**
  - These differences between model predicts and the actual predictions are called errors.
  - The goal of an analyst is not to eliminate errors but to reduce them.
  - There is always a tradeoff between how low you can get errors to be.
  - The prediction errors are usually known as **Bias and Variance**.
  - Aim of data scientist is to reduce these errors in order to get more accurate results for a particular dataset.
  - Let us understand bias and variance, Bias-variance trade-off, Underfitting and Overfitting.
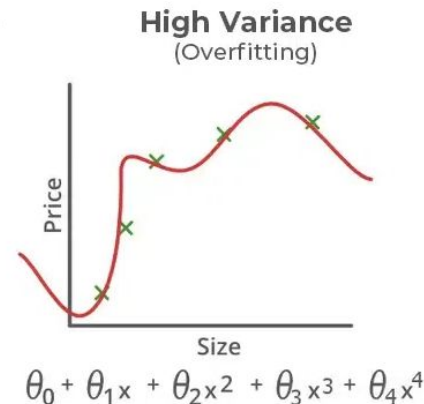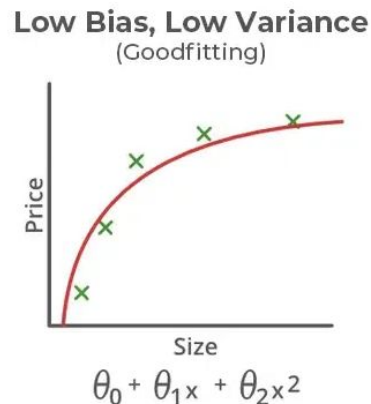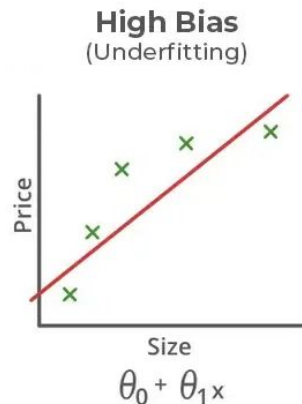
Errors in Machine Learning

- Error is a measure of how accurately algorithm can make predictions for seen or unseen dataset.
- We choose the ML model to reduce error and performs best for a particular dataset.
- Two main **types of errors** present in any ML model
- Bias and Variance in Machine Learning
- Reducible Errors
  - These errors can be reduced to improve the model accuracy.
  - Such errors can further be classified into **bias and Variance**.
- Irreducible Errors
  - These errors will always be present in the model due to unknown variables.
- Bias and Variance in Machine Learning
- However, achieving the balance between Bias and Variance can be challenging.
- Two common issues that affect model accuracy are overfitting and underfitting.
- These problems are major contributors to poor performance in ML models.

Bias and Variance in Machine Learning

- Bias
- Error due to very simple ML model which doesn't learn enough details from data.
  - High bias make the model easier to train but fails to captur the underlying complexities of data.
  - High bias typically leads to underfitting;
  - i.e model performs poorly on both training and testing data as it fails to learn enough from data.
- Variance
- Error due to perfect ML model which learns too much from the data, including random noise.
  - High-variance model learns not only the patterns but also the noise in the training data
  - High-variance leads to poor generalization on unseen data.
  - High variance typically leads to overfitting;
  - i.e model performs well on training data but poorly on testing data.

Bias–Variance Tradeoff

- The goal is to find an optimal balance where both bias and variance are minimized
- The relationship between bias and variance is often referred to as the bias–variance tradeoff.
- Tradeoff highlights the need for balance:
  - Increasing model complexity reduces bias but increases variance (risk of overfitting).
  - Simplifying the model reduces variance but increases bias (risk of underfitting).
- Example:
  - To predict the price of houses based on their size;
  - To draw a line or curve that best fits the data points on a graph.
  - Fitting line captures the trend in the data depends on the complexity of the used model.

High Bias (Underfitting)	Low Bias, Low Variance (Goodfitting)	High Variance (Overfitting)
$\theta_0 + \theta_1 x$	$\theta_0 + \theta_1 x + \theta_2 x^2$	$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

Overfitting

- Overfitting: The most common issues faced by Machine Learning engineers and data scientists.
- It occurs when ML model tries to cover all or more than the required data points in the given dataset.
- Training with large data leads to capturing noise and inaccurate data into the training data set.
- It negatively affects the performance of the model.
- Overfitting Example:
  - Training data sets such as 5000 mangoes, 1000 apples, and 1000 papayas.
  - Probability of identifying papaya as mangoes due to large biased data in the training data set; Hence prediction got negatively affected.
  - Overfitting caused by using non-linear methods of ML algorithms as they build non-realistic data models.
- We can overcome overfitting by using linear and parametric algorithms in the ML models.
- Example: Overfitting models
  - Students who memorize answers instead of understanding the topic.
  - These students do well in practice tests (training) but struggle in real exams (testing).
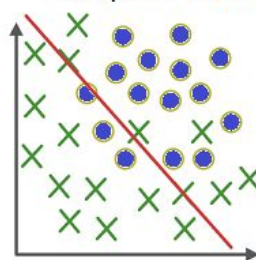
Overfitting

- Reasons for Overfitting:
  - High variance and low bias.
  - The model is too complex.
  - The size of the training data.
- Methods to reduce overfitting:
  - Increase training dataset.
  - Reduce model complexity by simplifying the model by selecting one with fewer parameters
  - Ridge Regularization and Lasso Regularization
  - Early stopping during the training phase
  - Reduce the noise
  - Reduce features
  - Reduce the number of attributes in training data.
  - Regularization: Controlling / constraining the model:
 - Ensemble Techniques

Underfitting

- Underfitting:
  - Underfitting: It is just the opposite of overfitting.
  - Underfitting occurs when ML model is unable to capture the basic underlying trend of the data.
  - It occurs due to training with fewer data and we try to build a linear model with non-linear data.
  - It provides incomplete and inaccurate data and destroys the accuracy of the ML model.
  - Underfitting occurs when our model is too simple to understand the base structure of the data.
- Example: Underfitting models
  - It is like students who don't study enough.
  - They don't do well in practice tests or real exams.
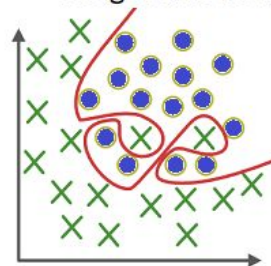 - The underfitting model has High bias and low variance.

Under-fitting
=> too simple
to explain variance

Okay-fitting
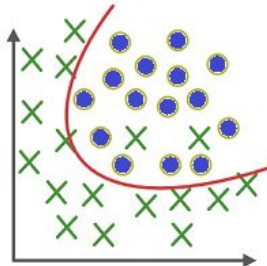
Over-fitting
=> Perfection:
too good to be true

Underfitting

- Reasons for Underfitting:
  - The model is too simple; So it may be not capable to represent the complexities in the data.
  - The input features are not the adequate to influence the target variable.
  - The size of the training dataset is not enough.
  - Excessive regularization are used to prevent the overfitting, it constraint model to capture the data well.
  - Features are not scaled.
- Methods to reduce Underfitting:
  - Increase model complexity
  - Increase the number of epochs to get better results.
  - Increase the training time of the model.
  - Increase the number of features.
  - Increased the quality of features
  - Remove noise from the data
  - Reduce the constraints

Model Evaluation

- Refer slide
  - vsat2k_ML_Ch1b Model Evaluation (Regularization) [PDF]

Confusion Matrix

- Confusion matrix shows # of correct and incorrect predictions made by
  - Classification model compared to the actual outcomes (target value) in the data.
- The matrix is NxN, where N is the number of target values (classes).
- Performance of such models is commonly evaluated using the data in the confusion matrix.

Dataset, N = 20 {4Class, 5Item}		Actual Class (Condition Given)					Y+N	Performance Measures
		Y			N			
Predicted Class (Outcome)	P	TP Correct Decision Prob = 1 – α	2	1	FP Type I Errors Prob = α	3	3	P: PPV = TP/(TP+FP) = 2 / (2+1) 0.667
	N	FN Type II Errors Prob = β	3	14	TN Correct Decision Prob = 1– β	17	17	NPV = TN/(TN+FN) = 3 / (3+14) 0.176
Total in Dataset	N+P	5			15		20	
Performance Measures		R: Sensitivity = TP/(TP+FN) = 2 / (2+3) 0.400			Specificity = TN/(TN+FP) = 14 / (14+1) 0.933			A=(TP+TN)/(TP+FP+TN+FN) = (2+14) / (2+1+3+14) 0.8

Confusion Matrix

Evaluation Metric	Formula	Value	Meaning
Precision P (PPV)	TP / (TP+FP)	0.667	Proportion of the retrieved documents that are correct
Recal R (Sensitivity)	TP / (TP+FN)	0.400	Proportion of the positives that the model retrieved
Accuracy (Recognition Rate)	(TP+TN) / Total	0.800	Proportion of total number of predictions that were correct
Error Rate	1 – Accuracy	0.200	Error with respect to recognition rate
Fall-out	FP / (FP+TN)	0.067	Proportion of non–relevant retrieved of all non–relevant.
F-Measure (F1 or F-Score)	2*(P*R)/(P+R)	0.500	Weighted harmonic mean of precision and recall.
Specificity (TN Recognition Rate)	TN / (TN+FP)	0.933	Proportion of actual –ve cases which are correctly identified.
NPV	TN / (TN+FN)	0.176	Proportion of –ve cases that were correctly identified.

Inference	Example	Meaning
Type I Error (α) or Level of Statistical Significance	For eg: α = 0.05; then the researcher has set 5% as the maximum chance of incorrectly rejecting the Ho.	Prob. of rejecting the Ho when it is actually true. (Falsely rejecting a Ho)
Type II Error (β)	For eg: β is set at 0.10, then the investigator has decided willing to accept a 10% chance of missing an association of a given effect size.	Prob. of failing to reject the Ho when it is actually false. (Falsely accepting a Ho)
Power($1 - \beta$)	If β is set at 0.10, then it represents a power of 0.90, i.e., a 90% chance of finding an association of that size.	Prob. of observing an effect in the sample (if any), of a specified effect size or greater exists in the population.
P Value	Ho is rejected in favor of the Ha if P value < α, the predetermined level of statistical significance.	Prob. of obtaining the study results by chance if the Ho is true.

Confusion Matrix

- Refer slide for more examples
  - Vsat IR – Confusion Matrix Complete [PDF]

References

Text books:

1. Ethem Alpaydin, "Introduction to Machine Learning , 4th Edition, The MIT Press, 2020.
2. Peter Harrington, "Machine Learning in Action", 1st Edition, Dreamtech Press, 2012."
3. Tom Mitchell, "Machine Learning", 1st Edition, McGraw Hill, 2017.
4. Andreas C, Müller and Sarah Guido, "Introduction to Machine Learning with Python: A Guide for Data Scientists", 1ed, O'reilly, 2016.
5. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", 1st Edition, MIT Press, 2012."

Reference Books:

6. Aurélien Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", 2nd Edition, Shroff/O'Reilly, 2019.
7. Witten Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal., "Data Mining: Practical machine learning tools and techniques", 1st Edition, Morgan Kaufmann, 2016.
8. Han, Kamber, "Data Mining Concepts and Techniques", 3rd Edition, Morgan Kaufmann, 2012.
9. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, "Foundations of Machine Learning", 1ed, MIT Press, 2012.
10. H. Dunham, "Data Mining: Introductory and Advanced Topics", 1st Edition, Pearson Education, 2006.

Thank You.

Satishkumar L. Varma