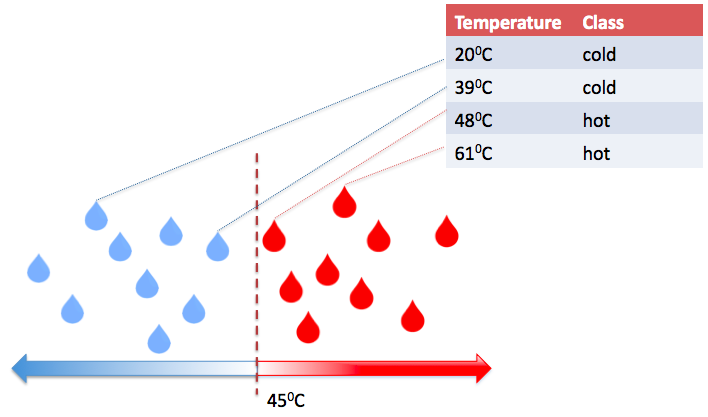
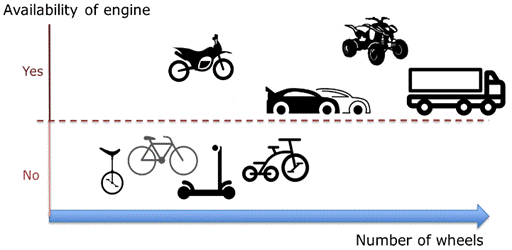
Feature- weather into good or bad we may need to rely on a number of characteristics including air temperature, humidity, wind strength and so on, rather than any single one. In machine learning terms, we call such characteristics features.

Classification-Discrete set of classes. Classifying things simply implies that we try to put them into clearly defined groups, classes or categories. In fact, we tend to classify all sorts of things all the time. Here are some examples:

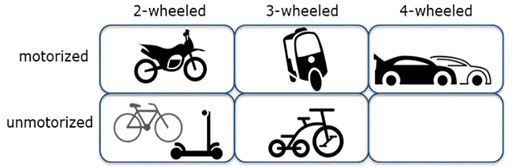


* based on our level of engagement and interest in a movie, we may classify it as interesting or boring;
* based on temperature, we classify water as cold or hot;
* based on the amount of sunshine, humidity, wind strength and air temperature, we classify the weather as good or bad;
* based on the number of wheels, we classify vehicles into unicycles, bicycles, tricycles, quadricycles, cars and so on;
* based on the availability of the engine, we may classify two-wheeled vehicles into bicycles and motorcycles.



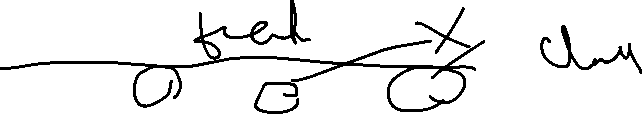
Multi Class Classification- Classification that implies more than two classes is called multi-class classification.

classify all vehicles based on the number of wheels, on size, or any other characteristics we will end up with multiple classes, for example, two-wheeled unmotorized vehicles, two-wheeled motorized vehicles, three-wheeled unmotorized vehicles



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | Sex | BP | Cholestrol | Drug |
|  |  |  |  | Drug A |
|  |  |  |  | Drug B |
|  |  |  |  | Drug C |

Decision tree



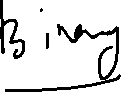
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | Sex | BP | Cholestrol | Drug |
| Young | M | High | High | Drug A |
| Middle Aged | F | High | High | Drug B |
| Senior | M | Med | Med | Drug B |
| Middle Aged | F | High | Med | Drug A |

Young M Med High ?

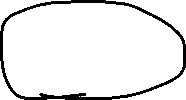
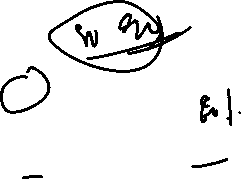


Recursive partitioning

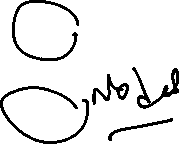
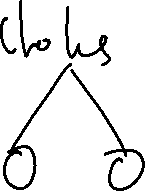
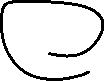
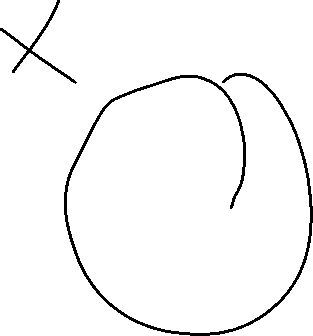
Pick up the best attribute to split the data



Age

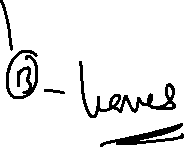
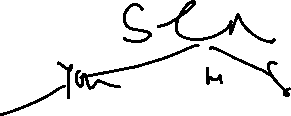
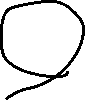
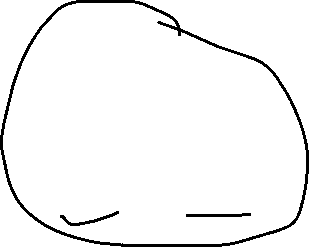


Young Middle Senior



B

Sex Cholestrol



B

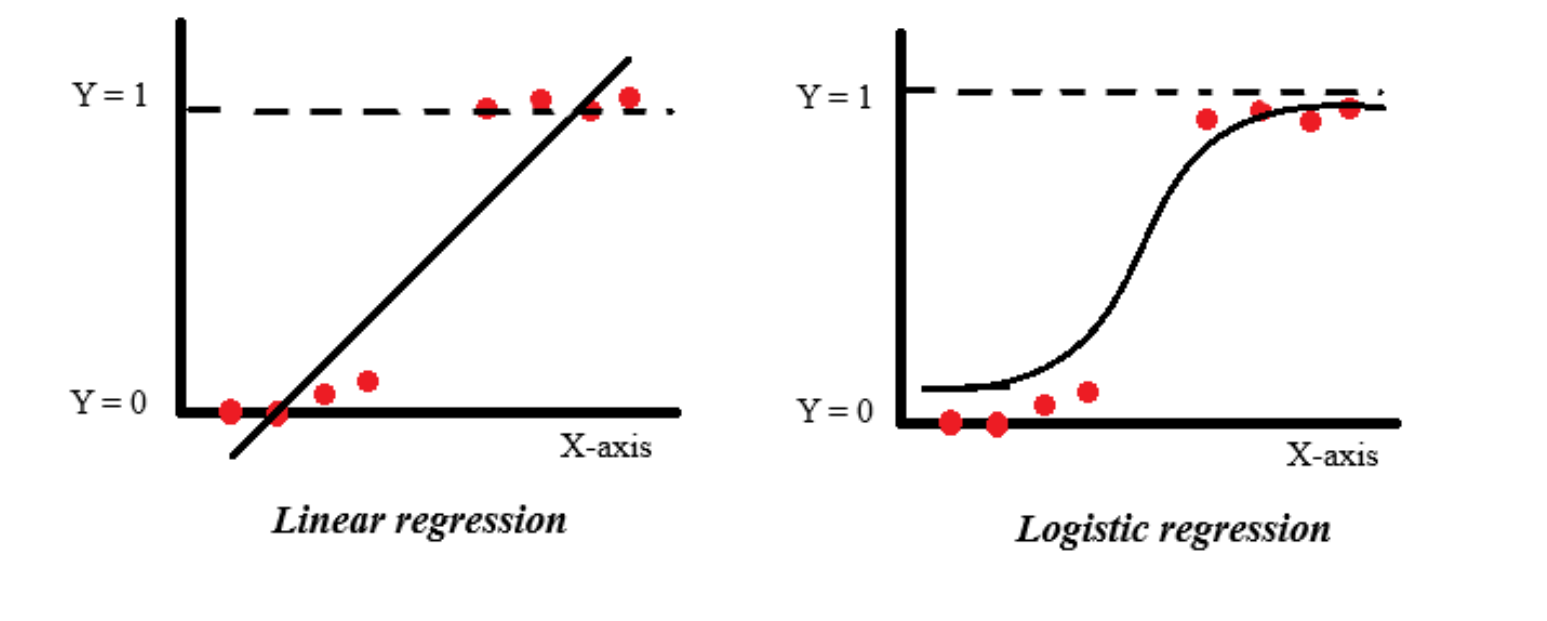
Logistic regression



1. In logistic regression, we use one or more independent variables such as tenure, age
2. and income to predict an outcome, such as churn, which we call a dependent variable,



1. representing whether or not customers will stop using the service.

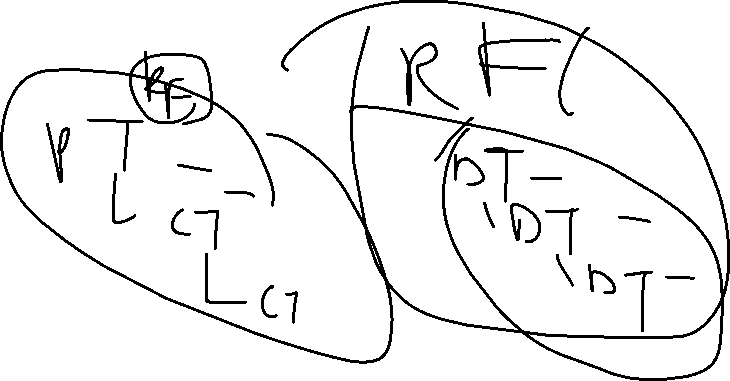


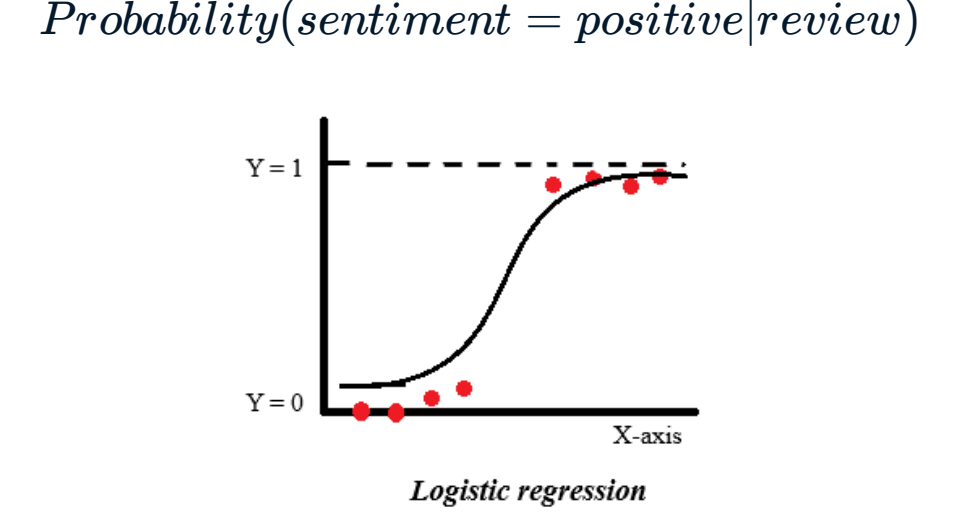


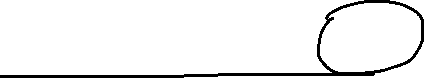
Sigmoid function = i/ 1+e^X



Value between 0 and 1





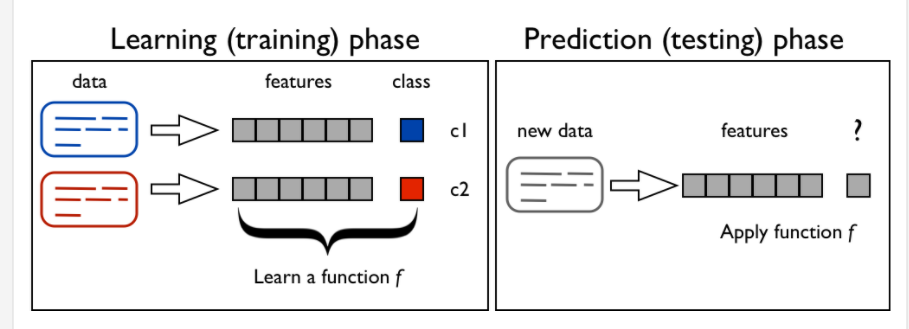


**Random Forest**

Its an ensemble algorithm.

In a random forest, an individual tree's vote impacts the final decision





Regression

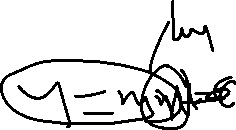
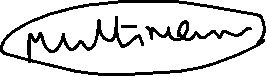
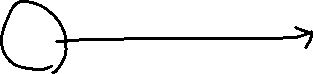
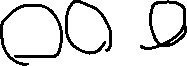
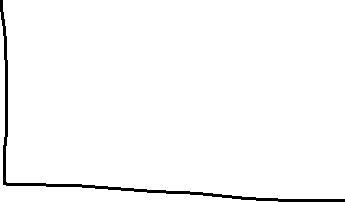
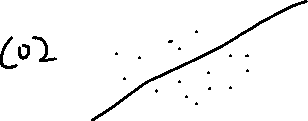
**regression is the process of predicting a continuous value.**



Continuous

|  |  |  |  |
| --- | --- | --- | --- |
| Engine Size | Cylinders | Fuel Consumption | CO2 Emissions |
| 2.4 | 4 | 8.5 | 196 |
| 2 | 4 | 9.6 | 221 |
| 1.4 | 4 | 5.9 | 135 |
| 3.5 | 6 | 11 | 255 |
| 2.7 | 6 | 10.6 | 244 |
|  |  |  |  |



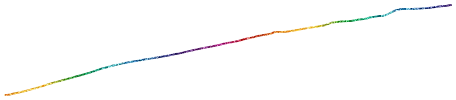


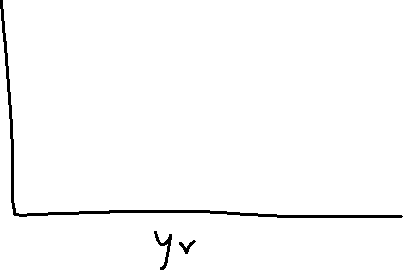
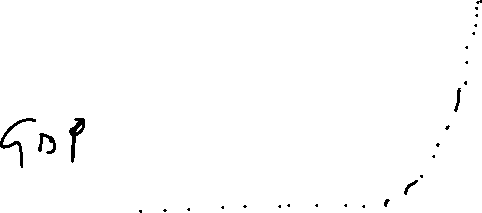
MSE=



Non Linear Regression

|  |  |
| --- | --- |
| Year | GDP |
| 1960 | 5.9 B |
| 1970 | 6.2 B |
|  |  |
| 2015 | 7.5 B |
| 2016 | 7.8 B |





Exponential or Logistical

y = e^ X

-Evaluation Metrics for NLP



| **Metric** | **Description** | **Applications** |
| --- | --- | --- |
| Accuracy [[48](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#footnote_2_30)] | Used when the output variable is categorical or discrete. It denotes the fraction of times the model makes correct predictions as compared to the total predictions it makes. | Mainly used in classification tasks, such as sentiment classification (multiclass), natural language inference (binary), paraphrase detection (binary), etc. |
| Precision [[48](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#footnote_2_30)] | Shows how precise or exact the model’s predictions are, i.e., given all the positive (the class we care about) cases, how many can the model classify correctly? | Used in various classification tasks, especially in cases where mistakes in a positive class are more costly than mistakes in a negative class, e.g., disease predictions in healthcare. |
| Recall [[48](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#footnote_2_30)] | Recall is complementary to precision. It captures how well the model can recall positive class, i.e., given all the positive predictions it makes, how many of them are indeed positive? | Used in classification tasks, especially where retrieving positive results is more important, e.g., e-commerce search and other information-retrieval tasks. |
| F1 score [[49](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#idm44947066748056)] | Combines precision and recall to give a single metric, which also captures the trade-off between precision and recall, i.e., completeness and exactness. F1 is defined as (2 × Precision × Recall) / (Precision + Recall). | Used simultaneously with accuracy in most of the classification tasks. It is also used in sequence-labeling tasks, such as entity extraction, retrieval-based questions answering, etc. |
| AUC [[48](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#footnote_2_30)] | Captures the count of positive predictions that are correct versus the count of positive predictions that are incorrect as we vary the threshold for prediction. | Used to measure the quality of a model independent of the prediction threshold. It is used to find the optimal prediction threshold for a classification task. |
| MRR (mean reciprocal rank) [[50](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#idm44947066738776)] | Used to evaluate the responses retrieved given their probability of correctness. It is the mean of the reciprocal of the ranks of the retrieved results. | Used heavily in all information-retrieval tasks, including article search, e-commerce search, etc. |
| MAP (mean average precision) [[51](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#idm44947066734168)] | Used in ranked retrieval results, like MRR. It calculates the mean precision across each retrieved result. | Used in information-retrieval tasks. |
| RMSE (root mean squared error) [[48](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#footnote_2_30)] | Captures a model’s performance in a real-value prediction task. Calculates the square root of the mean of the squared errors for each data point. | Used in conjunction with MAPE in the case of regression problems, from temperature prediction to stock market price prediction. |
| MAPE (mean absolute percentage error) [[52](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#idm44947066725304)] | Used when the output variable is a continuous variable. It is the average of absolute percentage error for each data point. | Used to test the performance of a regression model. It is often used in conjunction with RMSE. |
| BLEU (bilingual evaluation understudy) [[53](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#idm44947066720280)] | Captures the amount of n-gram overlap between the output sentence and the reference ground truth sentence. It has many variants. | Mainly used in machine-translation tasks. Recently adapted to other text-generation tasks, such as paraphrase generation and text summarization. |
| METEOR [[54](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#idm44947066715736)] | A precision-based metric to measure the quality of text generated. It fixes some of the drawbacks of BLEU, such as exact word matching while calculating precision. METEOR allows synonyms and stemmed words to be matched with the reference word. | Mainly used in machine translation. |
| ROUGE [[55](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#idm44947066697656)] | Another metric to compare quality of generated text with respect to a reference text. As opposed to BLEU, it measures recall. | Since it measures recall, it’s mainly used for summarization tasks where it’s important to evaluate how many words a model can recall. |
| Perplexity [[56](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#footnote_2_19)] | A probabilistic measure that captures how confused an NLP model is. It’s derived from the cross-entropy in a next word prediction task. The exact definition can be found at [[56](https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/ch02.html#footnote_2_19)]. | Used to evaluate language models. It can also be used in language-generation tasks, such as dialog generation. |