Essentials of Machine Learning Algorithms (with Python and R Codes)

BUSINESS ANALYTICS (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/BUSINESS-ANALYTICS/) PYTHON (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/PYTHON-2/) R (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/R/)

Analytics Vidhya
Learn Everything About Analytics

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

Google's self-driving cars and robots get a lot of press, but the company's real future is in machine learning, the technology that powers Google Search and allows the company to get smarter and more personal.

– Eric Schmidt (Google Chairman)

We are probably living in the most defining period of human history. The period when computing moved from large mainframes to PCs to cloud. But what makes it defining is not what has happened, but what is coming our way in years to come.

What makes this period exciting for some one like me is the democratization of the tools and techniques, which followed the boost in computing. Today, as a data scientist, I can build data crunching machines with complex algorithms for a few dollors per hour. But, reaching here wasn't easy! I had my dark

days and nights

The idea behind creating this guide is to simplify the journey of aspiring data scientists and machine learning enthusiasts across the world. Through this guide, I will enable you to work on machine learning problems and gain from experience. **I am providing a high level understanding about various machine learning algorithms along with R & Python codes to run them. These should be sufficient to get your hands dirty.**

POPULAR POSTS

- A Complete Tutorial to Learn Data Science with Python from Scratch (https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/)
- Essentials of Machine Learning Algorithms (with Python and R Codes) (https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/)
- A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python) (https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/)
- 40 Interview Questions asked at Startups in Machine Learning / Data Science (https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/)
- 7 Types of Regression

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scient)

(https://www.analyticsvidhya.com/wp-content/uploads
/2015/08/Essentials-of-Machine-Learning-Algorithms.jpg)

I have deliberately skipped the statistics behind these techniques, as you don't need to understand them at the start.

So, if you are looking for statistical understanding of these algorithms, you should look elsewhere. But, if you are looking to equip yourself to start building machine learning project, you are in for a treat.

Broadly, there are 3 types of Machine Learning Algorithms..

1. Supervised Learning

How it works: This algorithm consist of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data.

Techniques you should know! (https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/)
A Complete Tutorial on Time Series Modeling in R (https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/)
Beginner's guide to Web Scraping in Python (using BeautifulSoup) (https://www.analyticsvidhya.com/blog/2015/10/beginner-guide-web-scraping-beautiful-soup-python/)
Complete guide to create a Time Series Forecast (with Codes in Python) (https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/)

Examples of Supervised Learning: Regression, Decision Tree (https://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/), Random Forest (https://www.analyticsvidhya.com/blog/2014/06/random-forest-machine-learning-simplified/), KNN, Logistic

(https://datahack.analyticsvidhya.com/contest/skilltest-python-for-data-scien)

do not have any target or ate. It is used for clustering which is widely used for segmenting customers in different groups for specific interventions. Examples of Unsupervised Learning: Apriori algorithm, K-means.

3. Reinforcement Learning:

How it works: Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process

List of Common Machine Learning Algorithms

Here is the list of commonly used machine learning algorithms. These algorithms can be applied to almost any data problem:

1. Linear Regression

cost of houses, number of

nuous variable(s). Here, we

ependent and dependent

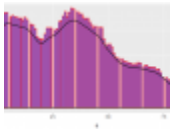
variables by fitting a best line. This best fit line is known as regression line and is represented by a linear equation Y= a *X + b.

The best way to understand linear regression is to relive this experience of childhood. Let us say, you ask a child in fifth grade to arrange people in his class by increasing order of weight, without asking them their weights! What do you think the child will do? He / she would likely look (visually analyze) at the height and build of people and arrange them using a combination of these visible parameters. This is linear regression in real life! The child has actually figured out that height and build would be correlated to the weight by a relationship, which looks like the equation above.

In this equation:

- Y – Dependent Variable
- a – Slope
- X – Independent variable
- b – Intercept

These coefficients a and b are derived based on minimizing the sum of squared difference of distance between data points and



regression line

Look at the below example. Here we have identified the best fit line having linear equation **y=0.2811x+13.9**. Now using this equation, we can find out the weight, knowing the height of a person.

Linear Regression is of mainly two types: Simple Linear Regression and Multiple Linear Regression. Simple Linear Regression is characterized by one independent variable. And, Multiple Linear Regression (as the name suggests) is characterized by multiple (more than 1) independent variables. While finding best fit line, you can fit a polynomial or curvilinear regression. And these are known as polynomial or curvilinear regression.

Python Code

```
braries like pandas, numpy...

from sklearn import linear_model

riable(s) and values mus

aining_datasets

t_datasets

ssion()

ing sets and check score

#Equation coefficient and Intercept

print('Coefficient: \n', linear.coef_)

print('Intercept: \n', linear.intercept_)

#Predict Output

predicted= linear.predict(x_test)
```

R Code

```
ets
onse variable(s) and values mus
t be numeric and numpy arrays
                    training_datasets
                    s_training_datasets
                    est_datasets

                                        score
                    x)
```


2. Logistic Regression

Don't get confused by its name! It is a classification not a regression algorithm. It is used to estimate discrete values (Binary values like 0/1, yes/no, true/false) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function (https://en.wikipedia.org/wiki/Logistic_function). Hence, it is also known as **logit regression**. Since, it predicts the probability, its output values lies between 0 and 1 (as expected).

Again, let us try and understand this through a simple example.

Let's say your friend gives you a puzzle to solve. There are only 2 outcome scenarios – either you solve it or you don't. Now imagine, that you are being given wide range of puzzles / quizzes in an attempt to understand which subjects you are good at. The outcome to this study would be something like this – if you are given a trignometry based tenth grade problem,

you are 70% likely to solve it. On the other hand, if it is grade probability of getting an answer is only 30%. This is what Logistic Regression provides you.

the outcome is modeled

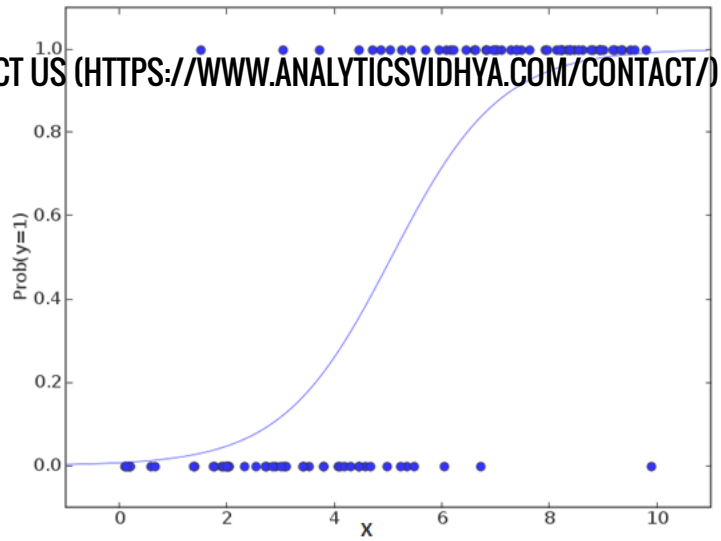
tor variables.

```
vent occurrence / probab
```

```
b2X2+b3X3....+bkXk
```

nce of the characteristic of interest. It chooses parameters that maximize the likelihood of observing the sample values rather than that minimize the sum of squared errors (like in ordinary regression).

Now, you may ask, why take a log? For the sake of simplicity, let's just say that this is one of the best mathematical way to replicate a step function. I can go in more details, but that will beat the purpose of this article.



(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Logistic_Regression.png)**Python Code**

(https://datahack.analyticsvidhya.com/contest/skilltest-python-for-data-scien)

```
#Assumed you have, X (predictor) and Y (target) for train
                                    ) of test_dataset

                                    ct

                                    ing sets and check score

                                    ef_)

                                    cept=)
```

R Code

```
x <- cbind(x_train,y_train)
# Train the model using the training sets and check score

logistic <- glm(y_train ~ ., data = x,family='binomial')
summary(logistic)

#Predict Output
predicted= predict(logistic,x_test)
```

Furthermore..

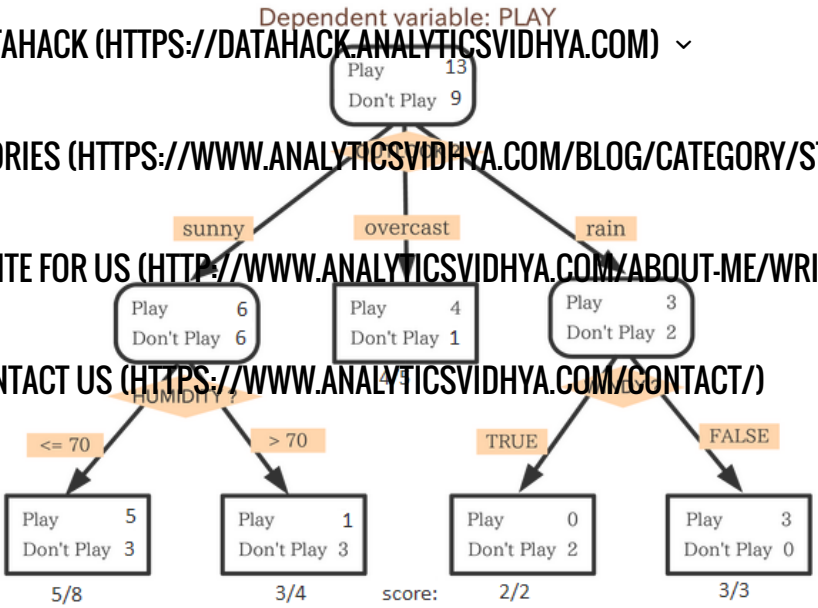
There are many different steps that could be tried in order to improve the model:

- including interaction terms
- removing features
- regularization techniques (https://www.analyticsvidhya.com/blog/2015/02/avoid-over-fitting-regularization/)

- using a non-linear model

3. Decision Tree

This algorithm is one of my favorite and I use it quite frequently. It is a type of supervised learning algorithm that is mostly used for classification problems. Surprisingly, it works for both categorical and continuous dependent variables. In this algorithm, we split the population into two or more homogeneous sets. This is done based on most significant attributes/ independent variables to make as distinct groups as possible. For more details, you can read: Decision Tree Simplified (http://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/).



(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/lkBzK.png)

source: statsexchange (http://stats.stackexchange.com)

In the image above, you can see that population is classified into four different groups based on multiple attributes to

identify 'if they will play or not'. To split the population into [illegible] groups, it uses various techniques like Gini, Information Gain, Chi-square, entropy.

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/download.jpg)

So, every time you split the room with a wall, you are trying to create 2 different populations within the same room. Decision trees work in very similar fashion by dividing a population in as different groups as possible.

More: Simplified Version of Decision Tree Algorithms (https://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/)

Python Code

```
braries like pandas, numpy...

from sklearn import tree

and Y (target) for train

) of test_dataset

er(criterion='gini') # f

m as g

by default it is gini

regression

ing sets and check score

predicted= model.predict(x_test)
```

R Code

```
library(rpart)

x <- cbind(x_train,y_train)

# grow tree

fit <- rpart(y_train ~ ., data = x,method="class")

summary(fit)

#Predict Output

predicted= predict(fit,x_test)
```

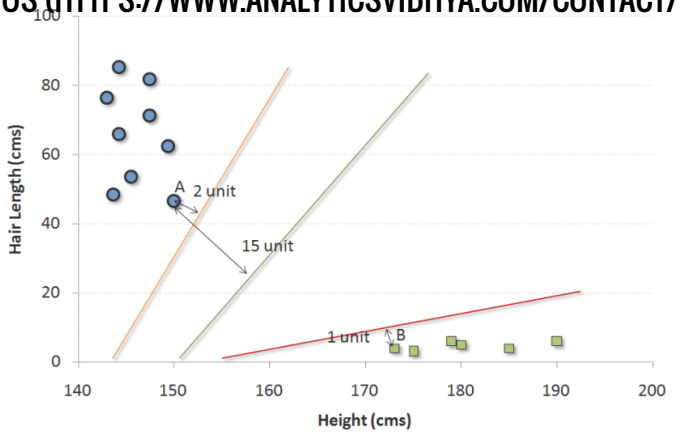
4. SVM (Support Vector Machine)

It is a classification method. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

For example, if we only had two features like Height and Hair (https://www.analyticsvidhya.com/wp-content/uploads/2015/08/SVM1.png) first plot these two variables in two dimensional space where each point has two co-ordinates (these co-ordinates are known as Support Vectors)

Now, we will find some *line* that splits the data between the two differently classified groups of data. This will be the line such that the distances from the closest point in each of the two groups will be farthest away.



(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/SVM2.png)

(https://datahack.analyticsvidhya.com/contest/skilltest-python-for-data-scien)

In the example shown above, the line which splits the data into two differently colored groups is the *black* line, since the two closest points are the farthest apart from the line. This line is a classifier. Then, depending on where the testing data is lying on either side of the line, we can classify what class we can classify

upport Vector Machine
log/2014/10/support-

JezzBall in n-dimensional

- You can draw lines / planes at any angles (rather than just horizontal or vertical as in classic game)
- The objective of the game is to segregate balls of different colors in different rooms.
- And the balls are not moving.

Python Code

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

```
#Assumed you have, X (predictor) and Y (target) for train
                                    ) of test_dataset
t
                                    us option associated wi
                                    ication. You can refer l
                                    he-machine-learning/), fo

                                    set and check score
```

R Code

```
library(e1071)
x <- cbind(x_train,y_train)

# Fitting model
fit <-svm(y_train ~ ., data = x)

summary(fit)
#Predict Output

predicted= predict(fit,x_test)
```

5. Naive Bayes

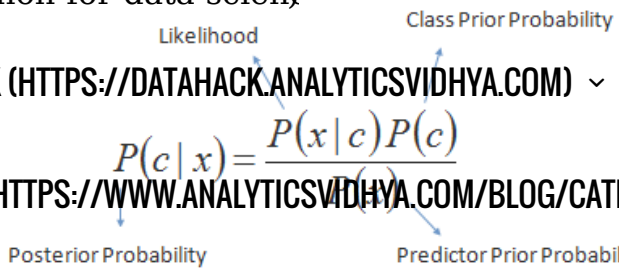
It is a classification technique based on Bayes' theorem (https://en.wikipedia.org/wiki/Bayes%27_theorem) with an assumption of independence between predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

other feature. For example, a fruit may be considered to be an [illegible] about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier would consider all of these properties to independently contribute to the probability

[illegible]uild and particularly useful [illegible] simplicity, Naive Bayes is [illegible]sophisticated classification

[illegible]culating posterior [illegible](x|c). Look at the equation below:



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Bayes_rule.png)

Here,

- *P*(*c*|*x*) is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- *P*(*c*) is the prior probability of *class*.
- *P*(*x*|*c*) is the likelihood which is the probability of *predictor* given *class*.
- *P*(*x*) is the prior probability of *predictor*.

Example: Let's understand it using an example. Below I have a

(https://datahack.analyticsvidhya.com/contest/skilltest-python-for-data-scien)

training data set of weather and corresponding target variable 'Play' (suggesting possibility of playing). Now, we need to classify whether players will play or not based on weather condition. Let's follow the below steps to perform it.

Step 1: Convert the data set to frequency table

Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Frequency table	
Weather	No
Overcast	
Rainy	3
Sunny	2
Grand Total	5

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
Grand Total	=5/14	=9/14		
	0.36	0.64		

Weather	Play
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Rainy	No

(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Bayes_41.png)

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Problem: Players will pay if weather is sunny, is this statement is correct?

We can solve it using above discussed method, so P(Yes | Sunny) = P(Sunny | Yes) * P(Yes) / P (Sunny)

Here we have P (Sunny |Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36, P(Yes)= 9/14 = 0.64

Now, P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

Naive Bayes uses a similar method to predict the probability of

different class based on various attributes. This algorithm is cation and with problems having multiple classes.


```
aussianNB
and Y (target) for train
) of test_dataset
= GaussianNB() #
multinomial classes like
modules/naive_bayes.html
# Train the model using the training sets and check scor
e
model.fit(x, y)
#Predict Output
predicted= model.predict(x_test)
```

R Code

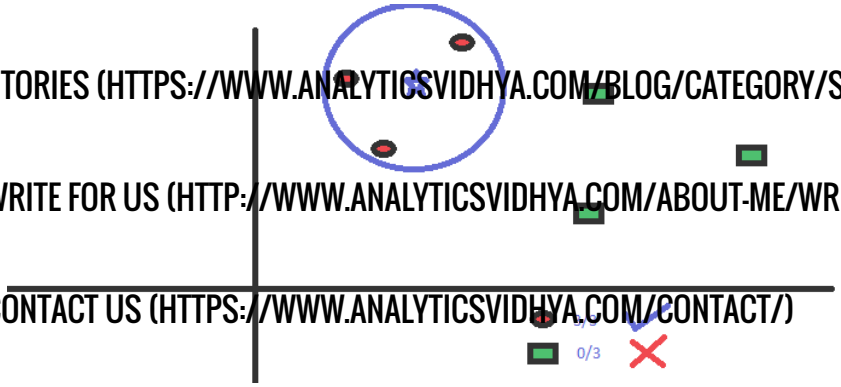
```
library(e1071)
x <- cbind(x_train,y_train)
# Fitting model
fit <-naiveBayes(y_train ~ ., data = x)
summary(fit)
#Predict Output
predicted= predict(fit,x_test)
```

6. KNN (K- Nearest Neighbors)

It can be used for both classification and regression problems. ...sification and regression problems. ...sed in classification problems in the industry. K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its ... neighbors. ...signed to the class is most ...neighbors measured by a ...e. Euclidean, Manhattan, ...ce. First three functions ...ne (Hamming) for ...e case is simply assigned ...At times, choosing K turns ...g KNN modeling.

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scienc)

Machine Learning Algorithms : K nearest neighbors : Simplified (http://Introduction to k-nearest neighbors : Simplified).



(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/KNN.png)

KNN can easily be mapped to our real lives. If you want to learn about a person, of whom you have no information, you might like to find out about his close friends and the circles he moves in and gain access to his/her information!

Things to consider before selecting KNN:

- KNN is computationally expensive
- Variables should be normalized else higher range variables can bias it
- Works on pre-processing stage more before going for KNN like outlier, noise removal


```
ighborsClassifier

and Y (target) for train
 for test_dataset

ject model

# Create KNeighbors
 # Train the model using the training sets and check score
 model.fit(X, y)

 #Predict Output
 predicted= model.predict(x_test)
```

R Code

```
library(knn)

x <- cbind(x_train,y_train)
# Fitting model

fit <-knn(y_train ~ ., data = x,k=5)

summary(fit)

#Predict Output

predicted= predict(fit,x_test)
```

7. K-Means

It is a type of unsupervised algorithm which solves the

clustering problem. Its procedure follows a simple and easy ... set through a certain number of clusters (assume k clusters). Data points inside a cluster are homogeneous and heterogeneous to peer groups.

...m ink blots? k means is ... ok at the shape and spread ... clusters / population are

(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/splatter_ink_blot_texture_by_maki_tak-d5p6zph.jpg)

How K-means forms cluster:

1. K-means picks k number of points for each cluster known as centroids.
2. Each data point forms a cluster with the closest centroids i.e. k clusters.
3. Finds the centroid of each cluster based on existing cluster members. Here we have new centroids.
4. As we have new centroids, repeat step 2 and 3. Find the closest distance for each data point from new centroids and get associated with new k-clusters. Repeat this process until convergence occurs i.e. centroids does not change.

How to determine value of K:

In K-means, we have clusters and each cluster has its own centroid. Sum of square of difference between centroid and the

(https://www.analyticsvidhya.com/wp-content/uploads
/2015/08/Kmenas.png)

Python Code


```
t KMeans

#Assumed you have, X (attributes) for training data set a
aset

ject model
dom_state=0)

ing sets and check score
```

```
fit <- kmeans(X, 3) # 5 cluster solution
```

8. Random Forest

Random Forest is a trademark term for an ensemble of decision trees. In Random Forest, we've collection of decision trees (so known as "Forest"). To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

Each tree is planted & grown as follows:

1. If the number of cases in the training set is N, then sample of N cases is taken at random but *with replacement*. This sample will be the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the

node. The value of m is held constant during the forest
(https://www.analyticsvidhya.)
e largest extent possible. There is no
pruning.

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

mparing with decision
would suggest you to

Simplified
m/blog/2014
ed/)

ndom Forest (Part 1)
m/blog/2014
/06/comparing-cart-random-forest-1/)

3. Comparing a Random Forest to a CART model (Part 2)
(https://www.analyticsvidhya.com/blog/2014
/06/comparing-random-forest-simple-cart-model/)
4. Tuning the parameters of your Random Forest model
(https://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/)

Python

```
...rt RandomForestClassifier
#Assumed you have, X (predictor) and Y (target) for train...
...) of test_dataset
...ing sets and check score
```



```
x <- cbind(x_train,y_train)
# Fitting model
fit <- randomForest(Species ~ ., x,ntree=500)
summary(fit)
#Predict Output
predicted= predict(fit,x_test)
```

9. Dimensionality Reduction Algorithms

In the last 4-5 years, there has been an exponential increase in data capturing at every possible stages. Corporates/ Government Agencies/ Research organisations are not only coming with new sources but also they are capturing data in great detail.

For example: E-commerce companies are capturing more details about customer like their demographics, web crawling history, what they like or dislike, purchase history, feedback and

many others to give them personalized attention more than (https://www.analyticsvidhya.) eper.

As a data scientist the data we are offered also consist of many ng good robust model but identify highly significant uch cases, dimensionality ong with various other dom Forest, PCA, Factor tion matrix, missing value

s, you can read "Beginners Reduction Techniques (https://www.analyticsvidhya.com/blog/2015/07/dimension-reduction-methods/)" /skilltest-python-for-data-scien)

Python Code

```
#Import Library
from sklearn import decomposition

#Assumed you have training and test data set as train and
 test

# Create PCA obeject pca= decomposition.PCA(n_components=
k) #default value of k =min(n_sample, n_features)

# For Factor analysis

#fa= decomposition.FactorAnalysis()

# Reduced the dimension of training dataset using PCA

train_reduced = pca.fit_transform(train)

#Reduced the dimension of test dataset

test_reduced = pca.transform(test)

#For more detail on this, please refer  this link (http:/

/scikit-learn.org/stable/modules/decomposition.html#decom

positions).
```

R Code

```
= TRUE)
train_reduced  <- predict(pca,train)
```


(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scienc)

AdaBoost

algorithms used when we
e a prediction with high
semble learning algorithm
several base estimators in
order to improve robustness over a single estimator. It combines multiple weak or average predictors to a build strong predictor. These boosting algorithms always work well in data science competitions like Kaggle, AV Hackathon, CrowdAnalytix.

More: Know about Gradient and AdaBoost in detail (https://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/)

Python Codes

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

```
rt GradientBoostingClassifier

#Assumed you have, X (predictor) and Y (target) for train
) of test_dataset

fier object
m_estimators=100, learn

_state=0)
ng.cons and print score
```

```
library(caret)

x <- cbind(x_train,y_train)
# Fitting model

fitControl <- trainControl( method = "repeatedcv", number
= 4, repeats = 4)

fit <- train(y ~ ., data = x, method = "gbm", trControl =
fitControl,verbose = FALSE)

predicted= predict(fit,x_test,type= "prob")[,2]
```

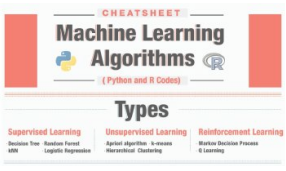
GradientBoostingClassifier and Random Forest are two different boosting tree classifier and often people ask about the difference between these two algorithms (http://discuss.analyticsvidhya.com/t/what-is-the-fundamental-difference-between-randomforest-and-gradient-boosting-algorithms/2341).

End Notes

By now, I am sure, you would have an idea of commonly used machine learning algorithms. My sole intention behind writing this article and providing the codes in R and Python is to get

you started right away. If you are keen to master machine
(https://www.analyticsvidhya.)
e up problems, develop a physical
understanding of the process, apply these codes and see the
fun!

RELATED



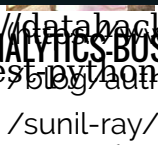
(https://www.analytic
/blog/2016/12/your-
review-analytics-

(https://www.analytic
a.com/
/blog/2015/09/full-
cheatsheet-machine-
ms/)
hon &
hon
g
nalytic
full-
hine-
ms/)
...Business Analytics"

(https://www.analytic
svidhya.com
/blog/2016/05/h2o-
data-table-build-
models-large-
data-sets/)
Use H2O and
data.table to build
models on large data
sets in R
(https://www.analytic
svidhya.com
/blog/2016/05/h2o-
data-table-build-
models-large-
data-sets/)
In "Machine Learning"

TAGS: C4.5 (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/C4-5/), CART
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/CART/), DECISION TREE
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/DECISION-TREE/), GBM
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/GBM/), K-MEANS
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/K-MEANS/), KNN
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/KNN/), LINEAR-REGRESSION
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/LINEAR-REGRESSION/), LOGISTIC REGRESSION
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/LOGISTIC-REGRESSION/), MACHINE LEARNING
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MACHINE-LEARNING/), NAIVE BAYES
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/NAIVE-BAYES/), NEURAL NETWORK
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/NEURAL-NETWORK/), RANDOM FOREST
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/RANDOM-FOREST/), REINFORCEMENT
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/REINFORCEMENT/), SUPERVISED LEARNING
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/SUPERVISED-LEARNING/), UNSUPERVISED
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/UNSUPERVISED/)

(https://datahack.analyticsvidhya.com/contest/skilltest-python-for-data-scien/)


(https://www.analyticsvidhya.com/blog/author/sunil-ray/)

Author
Sunil Ray
(https://www.analyticsvidhya.com/blog/author/sunil-ray/)

I am a Business Analytics and Intelligence professional with deep experience in the Indian Insurance industry. I have worked for various multi-national Insurance companies in last 7 years.

50 COMMENTS

Kuber says:
AUGUST 10, 2015 AT 11:59 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92380)
REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92380#RESPOND)

92380)

Analytics Vidhya (https://www.analyticsvidhya.com/)
Learn Everything About Analytics

n!! Thank you.

HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/) BLOG (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/)

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

COMMON-MACHINE-LEARNING-ALGORITHMS
JOBS (HTTPS://WWW.ANALYTICSVIDHYA.COM/JOBS/) CSVIDHYA.COM
ARNING-ALGORITHMS/#COMMENT-

TRAININGS (HTTPS://WWW.ANALYTICSVIDHYA.COM/TRAININGS/)
useful and excellent
ookmarked this page.

DISCUSS (HTTPS://DISCUSS.ANALYTICSVIDHYA.COM)

LEARNING PATHS (HTTPS://WWW.ANALYTICSVIDHYA.COM/LEARNING-PATHS-DATA-SCIENCE-BUSINESS-
COMMON-MACHINE-LEARNING-ALGORITHMS
/?REPLYTOCOM=92400#RESPOND) AUGUST 11, 2015 AT 4:50 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM
ANALYTICS-BUSINESS-INTELLIGENCE-BIG-DATA/)
/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-
92400)

DATAHACK (HTTPS://DATAHACK.ANALYTICSVIDHYA.COM)
Straight, rustraightand effective!
Thank you

STORIES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/STORIES/)

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS
/?REPLYTOCOM=92402#RESPOND) **Venugopal says:** AUGUST 11, 2015 AT 6:05 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM
/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-
92402)

WRITE FOR US (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/WRITE/)

CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)

Good Summary airticle

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS
/?REPLYTOCOM=92404#RESPOND) **Dr Venugopala Rao says:** AUGUST 11, 2015 AT 6:27 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM
/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-
92404)

Super Compilation...

(https://www.analyticsvidhya.)

HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/)
BLOG (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/)
JOBS (HTTPS://WWW.ANALYTICSVIDHYA.COM/JOBS/)
TRAININGS (HTTPS://WWW.ANALYTICSVIDHYA.COM/TRAININGS/)
DISCUSS (HTTPS://DISCUSS.ANALYTICSVIDHYA.COM)
LEARNING PATHS (HTTPS://WWW.ANALYTICSVIDHYA.COM/LEARNING-PATHS-DATA-SCIENCE-BUSINESS-ANALYTICS-BUSINESS-INTELLIGENCE-BIG-DATA/)
DATAHACK (HTTPS://DATAHACK.ANALYTICSVIDHYA.COM)
STORIES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/STORIES/)
WRITE FOR US (HTTP://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/WRITE/)
CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)

Kishor Basyal says:

...OG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS
...(HTTPS://WWW.ANALYTICSVIDHYA.COM
...ACHINE-LEARNING-ALGORITHMS/#COMMENT-92410)

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

...COMMON-MACHINE-LEARNING-ALGORITHMS
...WWW.ANALYTICSVIDHYA.COM
...RNING-ALGORITHMS/#COMMENT-

...this.

Tesfaye says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92416#RESPOND)

AUGUST 11, 2015 AT 10:30 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92416)

Thank you! Well presented article.

Tesfaye says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92417#RESPOND)

AUGUST 11, 2015 AT 10:31 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92417)

Thank you! Well presented.

Huzefa says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92436#RESPOND)

AUGUST 11, 2015 AT 3:53 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92436)

Hello,

Superb information in just one blog. Can anyone help me

to run the codes in R what should be replaced with "~"
help is appreciated

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scient

COMMON-MACHINE-LEARNING-ALGORITHMS
WWW.ANALYTICSVIDHYA.COM
ARNING-ALGORITHMS/#COMMENT-

can anyone help me
ould be replaced with "~"
reciated

Sunil Ray says:
AUGUST 13, 2015 AT 3:35 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92497)

Enjoyed the simplicity. Thanks for the effort.

Sunil Ray says:
AUGUST 14, 2015 AT 7:36 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92706)

Hi All,

Thanks for the comment ...

Dalila says:
AUGUST 14, 2015 AT 1:35 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92744)

Very good summary.

Thank!

e reason for taking the log(p)/(1-p) in
s to make the equation linear, I.e.,
easy to solve.

What's not the reason for taking the log? The underlying assumption in logistic regression is that the probability is governed by a step function whose argument is linear in the attributes. First of all the assumption of linearity or otherwise introduces bias. However, logistic regression being a parametric model some bias is inevitable. The reason to choose a linear relationship is not because its easy to solve but because a higher order polynomial introduces higher bias and one would not like to do so without good reason.

Now coming to the choice of log, it is just a convention. Basically, once we have decided to go with a linear model, in the case of one attribute we model the probability by

p(x) = f(ax+b)

such that p(-infinity)=0 and p(infinity)=0. It so
(https://www.analyticsvidhya.)
fies that this is satisfied by:

p(x) = exp(ax+b)/ (1 + exp(ax+b))

HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/) BLOG (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/)

JOBS (HTTPS://WWW.ANALYTICSVIDHYA.COM/JOBS/)

itten as

x+ b

TRAININGS (HTTPS://WWW.ANALYTICSVIDHYA.COM/TRAININGS/)

ay be useful to talk about

should ask is why we don't

ethod. The reason is that a

Bernoulli random variable

ate the probability according

od wrt Bernoulli process.

DISCUSS (HTTPS://DISCUSS.ANALYTICSVIDHYA.COM)

LEARNING PATHS (HTTPS://WWW.ANALYTICSVIDHYA.COM/LEARNING-PATHS-DATA-SCIENCE-BUSINESS-ANALYTICS-BUSINESS-INTELLIGENCE-BIG-DATA/)

n the assumption is that the

dependent variable' function are

distributed according to a normal distribution and the maximum likelihood estimate for a normal distribution amounts to the least square method. So deep down linear regression and logistic regression both use maximum likelihood estimates. Its just that they are max likelihoods according to different distributions.

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)
DATAHACK (HTTPS://DATAHACK.ANALYTICSVIDHYA.COM)
STORIES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/STORIES/)
WRITE FOR US (HTTP://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/WRITE/)
CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)

Statistics says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=93085#RESPOND)

AUGUST 19, 2015 AT 12:14 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-93085)

Nice summary!
@Huzefa: you shouldn't replace the "~" in the R code, it basically means "as a function of". You can also keep the "." right after, it stands for "all other variables in the dataset provided". If you want to be explicit, you can write y ~ x1 + x2 + … where x1, x2 .. are the names of the columns of your data.frame or data.table.
Further note on formula specification: by default R adds an intercept, so that y ~ x is equivalent to y ~ 1 + x, you can

remove it via y ~ 0 + x. Interactions are specified with
(https://www.analyticsvidhya.
ds the two variables) or : (which
tion term). y ~ x1 * x2 is equivalent to
y ~ x1 + x2 + x1 : x2.

is really helpful. Thanks!

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien

[illegible] (https://www.facebook.com/app_scoped_user_id/10153310687928141/) says:

SEPTEMBER 10, 2015 AT 7:48 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-94650)

I took the Stanford-Coursera ML class, but have not used it, and I found this to be an incredibly useful summary. I appreciate the real-world analogues, such as your mention of Jezzball. And showing the brief code snips is terrific.

Shankar Pandala (https://www.facebook.com/app_scoped_user_id/10153564059874654/) says:

SEPTEMBER 15, 2015 AT 12:09 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-95096)

This is very easy and helpful than any other courses I have completed.
simple. clear. To the point.

markpratley says:
...9 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/...MACHINE-LEARNING-ALGORITHMS/#COMMENT-96010)

...a scholar!

...COMMON-MACHINE-LEARNING-ALGORITHMS
...S://WWW.ANALYTICSVIDHYA.COM
...RNING-ALGORITHMS/#COMMENT-

...long with good examples and codes which is surely much helpful. Just, can you add Neural Network here in simple terms with example and code.

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=98659#RESPOND)

Sayan Putatunda says:
NOVEMBER 1, 2015 AT 7:00 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-98659)

Errata:- fit <- kmeans(X, 3) # 5 cluster solution
It`s a 3 cluster solution.

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=100598#RESPOND)

Baha says:
NOVEMBER 27, 2015 AT 1:13 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-100598)

Well done, Thank you!

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS

Benjamin says:

/?REPLYTOCOM=... NOVEMBER ..., 2015 AT 7:00 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM
/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-

This is a great resource overall and surely the product of ...

... is, your comment on Logistic
... regression is in fact wrong.
... variable ... between
... bability. it makes
... still an extra step that
... hold which is not the main
... a matter of fact it falls
... ized Linear Models ...
... code example.

I thought this was interesting to note so as not to forget that logistic regression output is richer than 0 or 1.

Thanks for the great article overall.

Bansari Shah says:
JANUARY 14, 2016 AT 6:27 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-103718)

Thank you.. reallu helpful article

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=103718#RESPOND)

---

ayushgg92 says:
JANUARY 22, 2016 AT 9:54 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-104348)

I wanted to know if I can use rattle instead of writing the R code explicitly

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=104348#RESPOND)

Debasis says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS

Nicolas Robles says:

...OG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS ...AM (HTTPS://WWW.ANALYTICSVIDHYA.COM ...ACHINE-LEARNING-ALGORITHMS/#COMMENT-105853)

...libraries...

...COMMON-MACHINE-LEARNING-ALGORITHMS ...S://WWW.ANALYTICSVIDHYA.COM ...RITHMS/#COMMENT-

(https://datahack.analyticsvidhya.com/co /skilltest-python-for-data-scien)

Col. Dan Sulzinger says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS /?REPLYTOCOM=106402#RESPOND)

MARCH 1, 2016 AT 1:21 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015 /08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-106402)

Good Article.

Pansy says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS /?REPLYTOCOM=106874#RESPOND)

MARCH 9, 2016 AT 5:04 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM /BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-106874)

I have to thank you for this informative summary. Really useful!

J (https://twitter.com/xuxeramos) says:

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS /?REPLYTOCOM=107024#RESPOND)

MARCH 10, 2016 AT 8:54 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM /BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-107024)

Somewhat irresponsible article since it does not mention

any measure of performance and only gives cooking
[illegible]standing what algorithm does what
it. Cooking recipes like these are the
ones that place people in Drew Conway's danger zone
[illegible]the-data-science-
[illegible]mmon-region-of-Hacking-
[illegible]dered-
[illegible]v.quora.com/In-the-
[illegible]why-is-the-common-region-
[illegible]antive-Expertise-considered-
[illegible]g programmers the worst
[illegible]requires another
recommend anyone wishing
[illegible]world not to jump into
[illegible]oper statistical background.
Otherwise you could end up like Google, Target,
[illegible]become a poster boy
for "The Big Flops of Big Data".

George says:
JUNE 17, 2016 AT 1:34 PM
(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-112323)

Do you have a better article? Please share...

Robin White (http://thedevmasters.com) says:
MARCH 15, 2016 AT 11:38 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-107434)

Great article. It really summarize some of the most important topics on machine learning.
But as asked above I would like to present thedevmasters.com as a company with a really good course to learn more depth about machine learning with great professors and a sense of community that is always

helping itself to continue learning even after the course

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

COMMON-MACHINE-LEARNING-ALGORITHMS
WWW.ANALYTICSVIDHYA.COM
ARNING-ALGORITHMS/#COMMENT-

is article to all my friends

COMMON-MACHINE-LEARNING-ALGORITHMS
WWW.ANALYTICSVIDHYA.COM
ARNING-ALGORITHMS/#COMMENT-

Very succinct description of some important algorithms.

Thanks. I'd like to point out a mistake in the SVM section. You say "where each point has two co-ordinates (these co-ordinates are known as Support Vectors)". This is not correct, the coordinates are just features. It's the points lying on the margin that are called the 'support vectors'. These are the points that 'support' the margin i.e. define it (as opposed to a weighted average of all points for instance.)

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=111356#RESPOND)

Isaac says:
MAY 24, 2016 AT 2:29 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-111356)

Thank you for this wonderful article...it's proven helpful.

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=112106#RESPOND)

Payal gour (http://www.excelclutchbrake.com/) says:
JUNE 11, 2016 AT 5:52 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-112106)

Thank you very much, A Very useful and excellent

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

COMMON-MACHINE-LEARNING-ALGORITHMS
W.ANALYTICSVIDHYA.COM/BLOG/2015
ORITHMS/#COMMENT-112309)

s of initial knowledge
hods can be fitted into a
might not be what you wish.
models, understand
diction really predicts.
is never ending
check for assumptions
validity.

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=112317#RESPOND)

Dhiraj Jain says:
JUNE 17, 2016 AT 10:24 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-112317)

The amazing article. I'm new in data analysis. It's very useful and easy to understand.

Thanks,

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=112851#RESPOND)

saharikannan says:
JUNE 20, 2016 AT 5:35 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-112851)

This is really good article, also if you would have explain about Anomaly dection algorithm that will really helpful for everyone to know , what and where to apply in machine learning....

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)

those who want to gain

COMMON-MACHINE-LEARNING-ALGORITHMS
//WWW.ANALYTICSVIDHYA.COM
RITHMS/#COMMENT-

Denis says:
SEPTEMBER 4, 2016 AT 10:53 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM
/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-
115567)

great summary,

Thank you

Faizan says:
SEPTEMBER 13, 2016 AT 9:17 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM
/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-
116024)

Thanks for the “jezzball” example. You made my day!

LEAVE A REPLY

Connect with:

(https://www.analyticsvidhya.com

/wp-login.php?action=wordpress_social_authenticate&
...ok&redirect_to=https%3A%2F
...1%2Fblog%2F2015%2F08%2Fcommon-
machine-learning-algorithms%2F)

ABOUT US

For those of you, who are wondering what is “Analytics Vidhya”, “Analytics” can be

LATEST POSTS

QUICK LINKS

Home (https://www.analyticsvidhya.com/)

defined as the science of

starts from capturing data and

(https://www.analyticsvidhya.com
/blog/2016/09/18-
free-exploratory-data-analysis-
tools-for-people-who-dont-
code-so-well/)

18 Free Exploratory Data Analysis Tools For People who don't code so well (https://www.analyticsvidhya.com/blog/2016/09/18-free-exploratory-data-analysis-tools-for-people-who-dont-code-so-well/)

MANISH SARASWAT , S…

(https://www.analyticsvidhya.com
/blog/2016/09/this-machine-
learning-project-on-imbalanced-
data-can-add-value-to-your-
resume/)

This Machine Learning Project on Imbalanced Data Can Add Value to Your Resume (https://www.analyticsvidhya.com/blog/2016/09/this-machine-learning-project-on-imbalanced-data-can-add-value-to-

to your resume/)

(https://datahack.analyticsvidhya.com/contest/skilltest-python-for-data-scien)

PySpark) (https://www.analyticsvidhya.com/blog/2016/09/comprehensive-introduction-to-apache-spark-rdds-dataframes-using-pyspark/)

ANKIT GUPTA , SEPTEM...

(https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/)

40 Interview Questions asked at Startups in Machine Learning / Data Science (https://www.anal

(https://datahack.analyticsvidhya.com/co
/skilltest-python-for-data-scien)