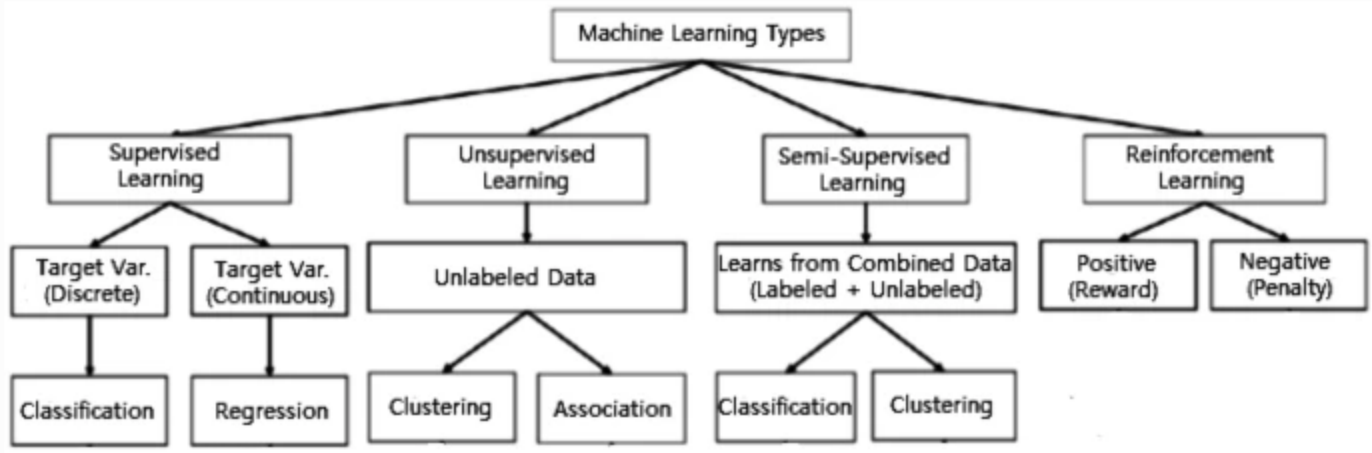
Fig. 2



Various types of machine learning techniques

Ref: https://link.springer.com/article/10.1007/s42979-021-00592-x

What are best practicies for ML?

1. Starting with an Interpretable Model

Keep the first model simple and get the infrastructure right far of any complications

2. Use Checkpoints:

A checkpoint is an intermediate dump of a model's internal state (parameters and hyperparameters),that will give high performance and less training time

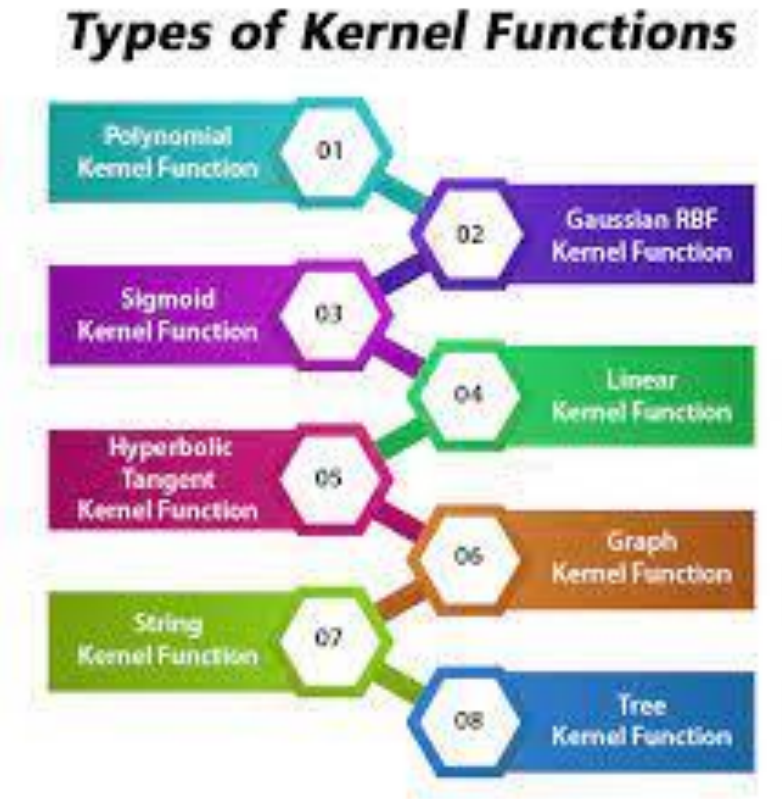
3. Performance over fancy metric

if there is some change that improves log loss but degrades the performance of the system, look for another feature.

4. Production Data to Training Data

store every new sample that comes from the serving model and then uses it for training.

Ref: https://rubikscode.net/2021/10/18/18-machine-learning-best-practices/



linear Kernel

It proves to be the best function when there are lots of features. The linear kernel is mostly preferred for text-classification problems as most of these kinds of classification problems can be linearly separated.

Linear kernel functions are faster than other functions.

$$F(x, xj) = sum(x.xj)$$

Here, **x, xj** represents the data you're trying to classify.

Polynomial Kernel

A General representation of linear kernel. It is not as preferred as other kernel functions as it is less efficient and accurate.

$$F(x, xj) = (x.xj+1)^d$$

Here '.' shows the **dot product** of both the values, and **d** denotes the degree.

F(x, xj) representing the **decision boundary** to separate the given classes.

Gaussian Radial Basis Function (RBF)

It is one of the most preferred and used kernel functions in svm. It is usually chosen for non-linear data. It helps to make proper separation when there is no prior knowledge of data.

$$F(x, xj) = exp(-gamma * ||x - xj||^2)$$

The value of gamma varies from **0 to 1**. You have to manually provide the value of gamma in the code. The most preferred value for **gamma is 0.1**.

Sigmoid Kernel

It is mostly preferred for neural networks. This kernel function is similar to a two-layer perceptron model of the neural network, which works as an activation function for neurons.

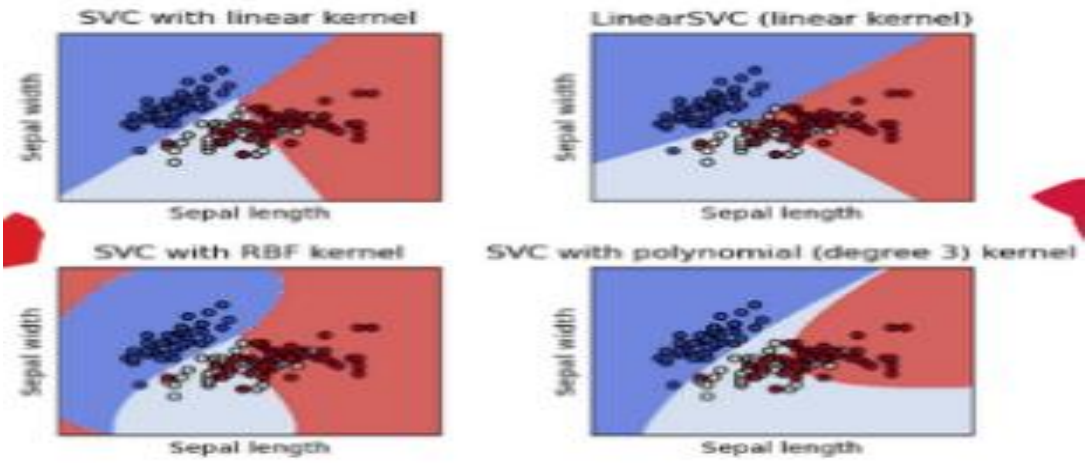
It can be shown as,

$$F(x, xj) = \tanh(\alpha xay + c)$$

Gaussian Kernel

It is a commonly used kernel. It is used when there is no prior knowledge of a given dataset.

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$



Ref : https://dataaspirant.com/svm-kernels/