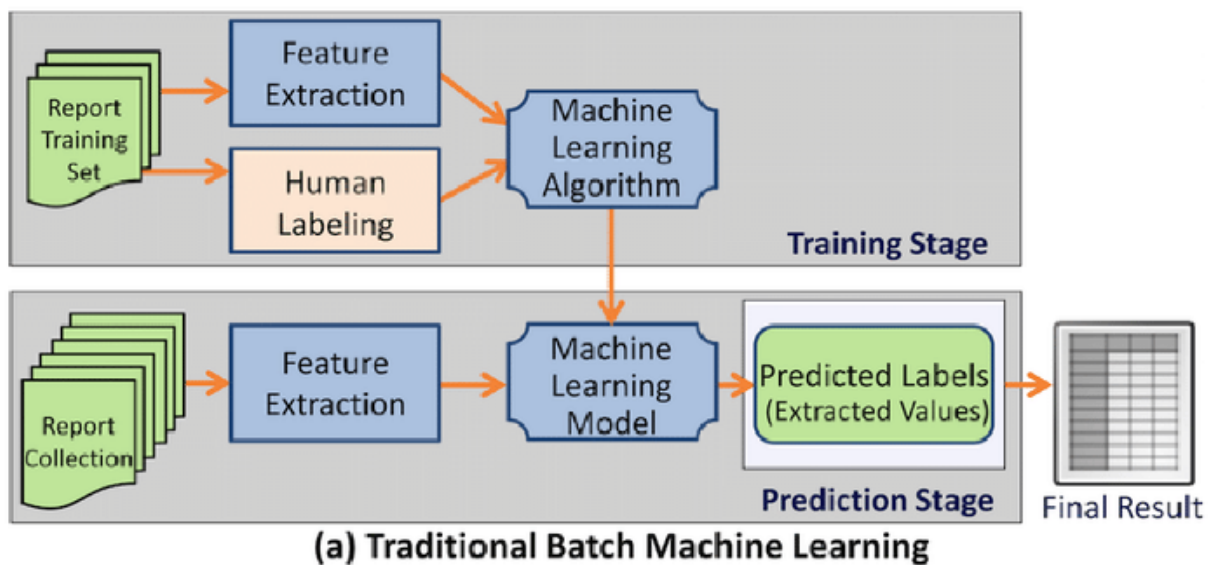


# ML-2

## Batch Machine Learning/ offline machine learning

in this type of machine learning , we train up our data at specific times in batches. It updates its parameters only when a new batch arrives. As they get trained in large batches they require high CPU , memo space , disk etc.



### Batch size

Now, the size of the batches are divided into 3 types:

#### Batch Mode

size of batch is equal to the total dataset

#### Mini-Batch Mode

size of the batch is greater than 1 but less than the total dataset.

#### Stochastic Mode

size of the batch is equal to one .

## Advantages

- efficiency is increased
- quality control is high
- production cost is less

## Disadvantages

- needs more time and resources
- they are monolithic in nature
- very complex to manage

# Online machine learning

Online machine learning, also known as incremental learning or real-time learning, is more dynamic as compared to the batch . here the model continuously learns from the data in real time .

Libraries used in python are: `riverml.xyz` and `vowpalwabbit`

## Advantages

- Adaptability
- Scalability
- Real time predictions

## Disadvantages

- Tricky to use
- less control over training
- lack of interpretability

**Note:**

if we face any problem in the online system , we use the anomaly detection system .

For , this we need to change the online machine learning to the batch machine learning . so, we rollback .

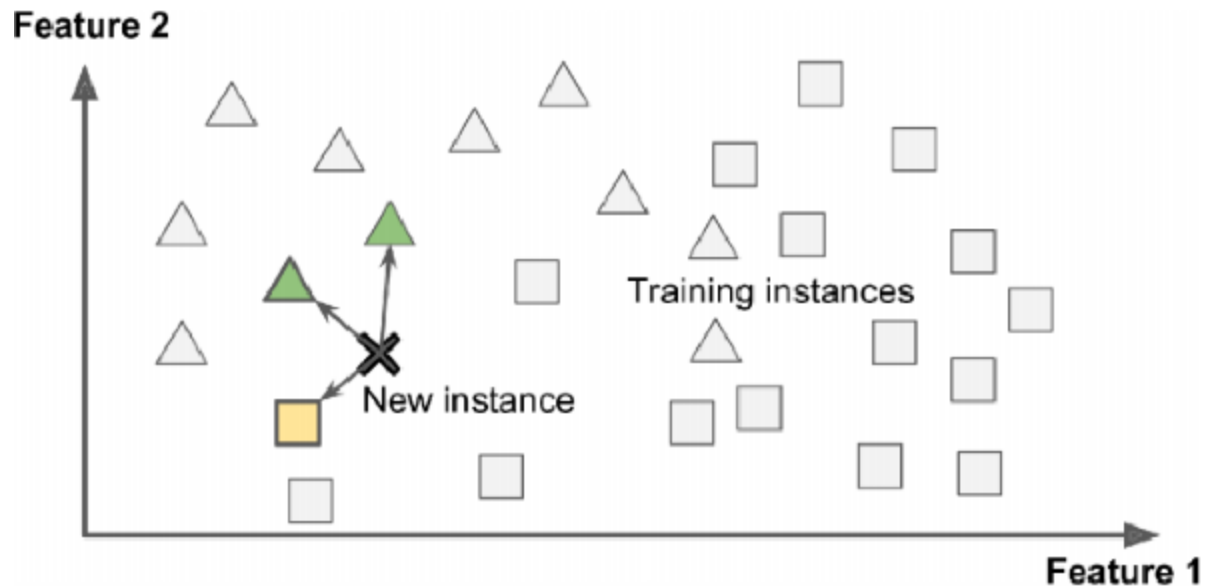
## Difference between online and offline machine learning

Offline Learning	Features	Online Learning
Less complex as model is constant	Complexity	Dynamic complexity as the model keeps evolving over time
Fewer computations, single time batch-based training	Computational Power	Continuous data ingestions result in consequent model refinement computations
Easier to implement	Use in Production	Difficult to implement and manage
Image Classification or anything related to Machine Learning - where data patterns remains constant without sudden concept drifts	Applications	Used in finance, economics, health where new data patterns are constantly emerging
Industry proven tools. E.g. Sci-kit, TensorFlow, Pytorch, Keras, Spark Mlib	Tools	Active research/New project tools: E.g. MOA, SAMOA, scikit-multiflow, streamDM

# Instance based machine learning

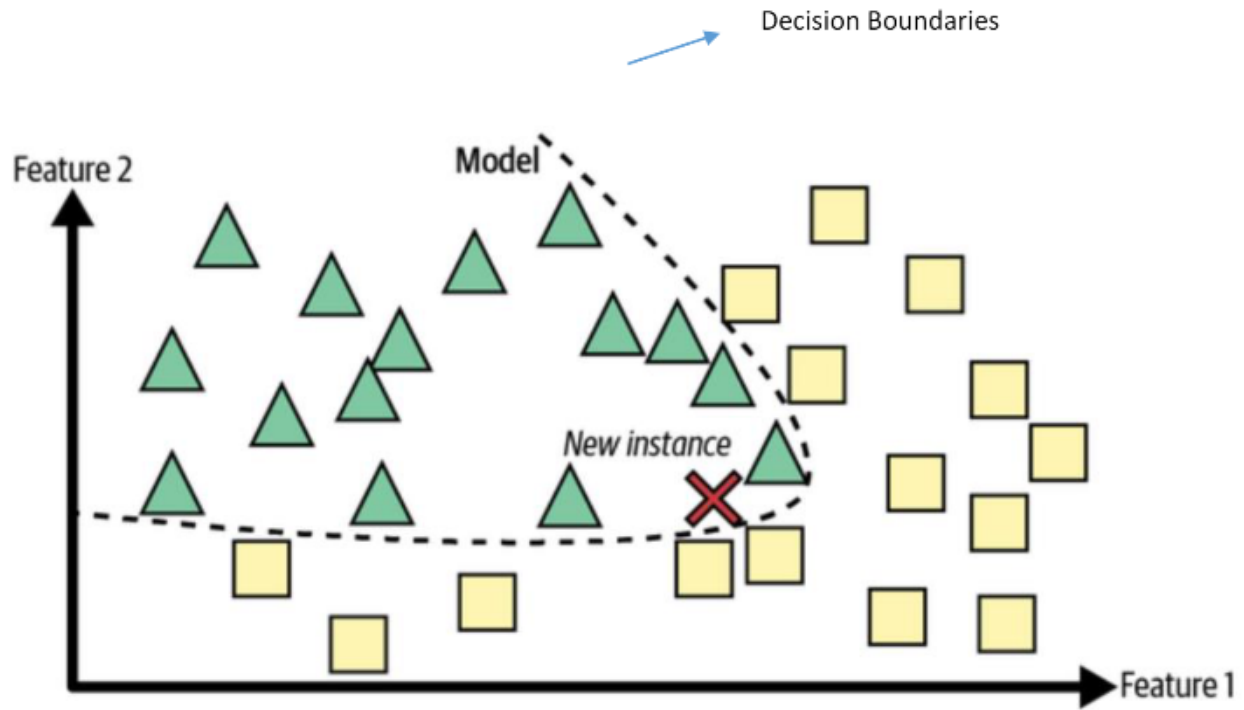
also known as lazy or memory based machine learning (because they delay the processing until the new instance is clarified). it makes hypotheses from the training instances.

time complexity depends upon the amount of data.



# Model based machine learning

also called structure based or eager learning. it involves the algos of linear , logistic and random forest. Model is explicitly created to represent the underlying structure of the data



It has faster processing speed and better generalization capabilities but it requires more time and effort to develop and tune the model for optimal model.

**Instance based**

**Model-based**

Prepare the data for model training	Prepare the data for model training. No difference here
Train model from training data to estimate model parameters i.e. discover patterns	Do not train model. Pattern discovery postponed until scoring query received
Store the model in suitable form	There is no model to store
Generalize the rules in form of model, even before scoring instance is seen	No generalization before scoring. Only generalize for each scoring instance individually as and when seen
Predict for unseen scoring instance using model	Predict for unseen scoring instance using training data directly
Can throw away input/training data after model training I	Input/training data must be kept since each query uses part or full set of training observations
Requires a known model form	May not have explicit model form
Storing models generally requires less storage	Storing training data generally requires more storage