subfields of ML

deeplearning

supportvector mchines etc

how ML works

problem statement => data collection => data cleaning => Data analysis and exploration => Data modeling => Optimization and deployment

applications od ML

spam detection

web advertisemnets

froud detection, social networking, vehicle driver assistance, speech recognition

web searches, etc

basic idea of an algorithm

input data(x) 🡺 set of rules(algorithm) 🡺 output response(y)eg

(1,5,10) 🡺 +1 🡺 (2,6,11)

Types of algorithms

* Regression algorithm
* Classification algorithm

Machine learning life cycle

1. Problem definition
   * Collaboration: working together with stake holders
   * Clarity: clearly write the objectives
   * Foundation: Establish a solid foundation for the machine
2. Data collection
   * Relevance
   * Quality: gather accurate data
   * Quantity: gather sufficient data
   * Diversity:
3. Data cleaning and pre processing
   * Data cleaning: data sorting
   * Data processing: data encoding
   * Data quality
4. Exploratory data analysis (EDA)
   * Exploration
   * Patterns and trends– qn what should be done when data is unbalanced
   * Insights: deleting useless columns
   * DECISION MAKING
5. Feature engineering and selection
   * Feature engineering
   * Feature selection
   * Domain expertise
   * Optimization
6. Model selection
   * Alignment
   * Complexity
   * Descision factors
   * Experimentation: experiment with different modles. Use tools like pycart
7. Model training
   * Training data
   * Iterative proceses
   * Optimization: fine tune the model
   * Validation: rigorous training to ensure accuracy to new unseen data
8. Evaluation and tuning
   * Evaluation metrics
   * Strength and weakness
   * Iterative improvement
   * Model robustness
9. Model deployment
   * Integration
   * Descision making
   * Practical solution
   * Continuous improvement.

Differences between supervised and unsupervised learning

Types of supervised learning,- regression(continuous/numerical) and classification(cartegorical)

Types of unsupervised learning,- clastering, trend/topic detection, dimensionality reduction

Types of classification:

* Binary classification
* Multiclass classification: have more than two classes
* Multilevel classification
* Imbalanced classification: number of classes in each class is unequally distributed

Classification algorithms

* Logistic regression
* Knearest neighbours
* Descision trees
* Support vector machines
* Naïve bayes
* Gradient bosting
* Neural networks

Setting up a supervised ML expt

* Gather
* Split into training and testing
* Train the model and tune
* Merge training and tunning to train a final model
* Run final model on the testing partition and evaluate.

How much training is enough?

* + Overfitting
  + Underfitting
  + Generalization
  + Regularization

Evaluating model performance

* Evaluation metrics(sensitivity, specificity, accuracy, F-Measure
* Error metrics(confusion matrix, accuracy(No. correct decisions)/(wrong+correct decisions)
* Recall/sensitivity

KNeighbours classifier snntax.

* + From sklearn.neighbours imort KNeighboursClassifier
  + KNN

Logistic regression syntax