**MACHINE LEARNING**çizgi film, Animasyon, ekran görüntüsü, kırpıntı çizim içeren bir resim

Açıklama otomatik olarak oluşturuldumetin, daire, logo, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Tom Michell – one of the founders of ML

**Definition**: If it’s performance on T, as measured by P, improves with E, then it is learning.

* Task 🡪 recommending a song
* Experience (data)
* Performance measure

**Learning Approaches**

1. Supervised Learning: learning with **labeled** data (training set). -> **classification** and **regression**
2. Unsupervised Learning: discovering patterns in **unlabeled** data. -> **clustering**
3. Semi-Supervised Learning: labeled and unlabeled data.
4. Reinforcement Learning: learning based on **feedback** (no dataset)

**metin, ekran görüntüsü, yazı tipi, grafik tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**Supervised Learning**

\* uses **classification** and **regression** algorithms (\*binary Linear, Nonlinear classifier)

Regression: real value. predict salary from age / credit amount

Classification: discrete(-1,1). predict class of data / receive credit or not

\* datayı önceden bilinen classlara ayırıyor. \* 2 class=binary classifier, 3=multiclass

– K-nearest neighbours, Support vector machines, Neural networks, Decision trees, Naive bayes

**K-Nearest Neighbours (K-NN) -classification**

-similarity between nighbours -value of k is very important

Uses euclidian distance to find nearest neighbours

Euclidian distance küçükse, daha benzerdir

Approachs used euclidian distance: Majority(4 out of 5), Weighted distance

Do not build a model -if we have categorical information, we can use encoding schemas

**Linear Regression**

Just works in binary classification – 2 choice (not works in multiclass) – draw line

Normal Equation – no iteration – no specify a convergence – only if x is invertible -slow

Gradient descent

In loss function(bowl), start at some point and gradually descent the error values,

so we obtain 0 slope point=global minimum point=convergence

“Kayıp fonksiyonunda bir noktadan başla ve hata değerlerini kademeli azalt, 0 eğim noktası = minimum

Need to choose learning rate alfa

**Logistic Regression** (actually it is CLASSIFICATION method)

sigmoid function –if there are many local optimas, gradient descent cant find global optimum

convex function

**Tree Classifiers (Decision Trees)**

There is a RULE, no assumption about linearity(linear-logistic)

No need to 2d data, no need to turn into numerical features

Build the tree until obtain pure nodes(at the leaves)

Just uses categorical features

Not builds function

\*C4.5 entropy – select next attribute to split on – pure=min homogenity/impurity, entropy(s)=0

Information gain – expected reduction in entropy –> ilk entropy-o seviyedeki tüm entropyler

If information gain=0, then we obtain nothing by applying this feature as splitting criteria, at this level

Highest gain seçilir

Training and Testing

**Overfitting** – fits all training data – makine ezberlemesi - train setini çok iyi öğrenmiş, test verilerini iyi sınıflandıramıyor

**Underfitting** – training seti bile classify edemiyor

K fold cross validation – 20 veri var, 2(k) gruba ayırıyorsun, 1 tanesi test 9u train

Prunning strategies:

To get suitable tree sizes and avoid overfitting,

-first grow complex tree, then start prunning

-use validation set -rule post prunning:turn the decision tree into set of rules

CART – top down algorithm, binary splits, gini index instead of entropy

**Naive Bayes Classifier**

Practical – easy to calculate

Useful for huge amount of data

Naive beacuse the independence assumption(all features are independent) – not realistic

Training data, feature vector, discrete label. d features and n examples

Perceptron Algorithm perceptron=single neuron

-linear classification -2 groups -simplest classification and neural network

-works only in linearly seperable data

**-iteratively update weights in order to find hyperplane that linearly seperates input data**

-and, or, nand, nor, not can be represented ; XOR cannot represented(nonlinear)

Perceptron convergence theorem: any linear function can be learned using this algorithm in a finite number of iterations

Uses threshold(step) function, undifferentiable so not suitable for gradient descent

Sigmoid function is differentiable

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Neural Networks** -handwritten character recognition

Uses the ability of perceptrons to represent elementary functions and

combine them in a network layers of elemantry questions

Propogation

Convergence: small changes i weigts

**Unsupervised Learning**

\* uses **clustering** algorithms (cannot classify)

– K means, Gaussian mixtures, Hierarchical clus., Spectral clus.

K-means

-easy to implement

Cons: need to know K, curse of dimension exponential

Feature engineering: all futures are not useful, they eliminate them.