**GREEDY**

-greedy e determinist: ofera acelasi raspuns la 2 rulari diferite, avand aceleasi date de intrare

-poate identifica o solutie a problemei, dar nu o identifica intotdeauna

-foloseste DFS

**OPERATORI**

-operatori de variatie (actioneaza la nivelul individului):

\*de incrucisare

\*de mutatie

1. uniforma: fiecare gena are sanse egale de a suferi o mutatie

2. neuniforma: prob de mutatie variaza pe parcurs(in functie de iteratie sau starea algoritmului evolutiv)

\*de recombinare

-este un operator exploativ pentru ca se concentreaza pe informatia genetica deja existenta in populatie

-operatori de selectie (actioneaza la nivelul populatiei)

\*bazat pe ruleta (pe baza de probabilitati)

\*bazat pe ranguri (in functie de fitness-ul fiecarui individ)

\*selectarea turnirului (alegere random -> best child)

\*selectare elitista

- Selecţia pentru supravieţuire

(μ,λ) sau (μ+λ)

**ALTELE DESPRE ALGORITMI EVOLUTIVI**

-tipuri: steady(aici concureaza parintii cu copiii pt supravietuire), generational (aici toata populatia e inlocuita de descendenti, incrucisarea si mutatia se fac pe fiecare generatie)

-steady modifica cel mai slab cromozom

-la comis voiajor, numarul de gene este egal cu numarul de orase

-un algoritm evolutive NU este determinist

-populatia initiala e aleasa alator, dar pot exista si mutatii

-modelul de populatie generational: intrega populatie se schimba intre generatii

-dezavantaje: over-fitting, nu exista garantia ca va gasi maxim global, timp de executie lung

-poate rezolva probleme de optimizare sau modelare

**FORMULE**

**-**A picture containing text, screenshot, font

Description automatically generated

-recall (real data R): \*TP/(TP+FN) unde TP=true positive si FN=false negative

\*se refera la proportia tuturor rezultatelor relevante

-precizie(predicted data P): TP/(TP+FP)

-scor: 2PR/(P+R)

-acuratete: total corecte din tot setul de prezis (nu pe o clasa anume)

-precizia evalueaza performanta in contextual unei probleme de clasificare

-daca clasificatorul A are precizia si recall-ul ai bune decat clasificatorul B, atunci A este probabil mai bun decat B (separate nu tare putem fi siguri)

-matricea de confuzie ofera date despre toate cele 3 (acuratete, precizie si recall) pentru a pute a fi calculate (ex pt ala cu florile, cate au fost prezise ca fiind correct in clasa respective erau pe diagonala principala si retul au fost prezise gresit in celelalte 2 clase)

-eroare medie patratica (mean square error)

MSE = (1/n) \* Σ(y - y\_pred)^2

unde:

n reprezintă numărul de exemple în setul de date,

y reprezintă valoarea reală a variabilei de ieșire,

y\_pred reprezintă valoarea prezisă a variabilei de ieșire.

**REGRESIE**

-INTERCEPTUL NU E ATRIBUT

-univariata: avem o singura variabila independenta si o singura variabila dependenta

-prezicerea unui output pe baza unor date anterioare (valoare numerica continua)

-metode:

\*metoda celor mai mici patrate

\*Ridge regression (aia cu determinantul)

-coeficient de corelatie: nr intre -1 si 1, -1 insemnand invers proportionalitate perfecta si 1 insemnand proportionalitate perfecta

-eroarea medie patratica

\*pentru fiecare observație în setul de date, preziceți valoarea corespunzătoare utilizând modelul de regresie

\*calc diferența între valoarea prezisă și valoarea reală pentru fiecare observație

\*ridică la pătrat fiecare diferență calculată.

\*calculează media pătratelor diferențelor pentru toate observațiile.

-ne trebuie variabile continua, nu nominale

-aici avem regresie liniara: ax+b

A picture containing text, screenshot, font

Description automatically generated

**CLASIFICARE**

-prezicerea clasei de apartenenta aunui obiect, pe baza caracteristicilor sale

-se foloseste k-means pt problemele cu invatare automata

\*poate fi evaluat folosind indecsi de evaluare interna

\*metrici pt compararea exemplelor folosind: dist euclidiana, dist manhattan

-gradient descendent e folosit tot aici

-in prob de clasificare, precizia se poate folosi ca masura care evalueaza performanta

-regresia logistica pt invatare supervizata

\*se foloseste in alg de invatare automata cu date de iesire de tip discret (numarabili)

\*A picture containing font, line, text, white

Description automatically generated

\*se poate evalua folosind acuratetea, rapelul, NU MEAN SQUARE ERROR

\*folosind SIGMOID

\*are codomeniul (0,1)

\*folosita siplu pt binar si one-vs-all pt multiclass

**GRADIENT DESCENDENT**

-minimizarea/optimizarea erorii obtinute pt un model de predictie

-un batch mic functioneaza mai bine pentru foarte multe exemple

-optimizeaza prin minimizare o functie de COST asociata unui algoritm de ml

-exista gradient descendent mini-batch:

\*Eroarea se calculează pentru fiecare exemplu de antrenament dintr-un mini-batch

\*fiecare exemplu de antrenament dintr-un mini-batch

**NORMALIZARE**

-min-max: (val - minim) / (maxim - minim)

**RANDOM**

-rucsac:

-c=(1,0,1,1,0)

-fenotip: obiectele 1,3,4 sunt incluse in rucsac

-genotip: fix repr asta (1,0,1,1,0)

-distanta euclidiana: sqrt( (x2-x1)^2 + (y2-y1)^2 )

-indexul david-bouldin

Indexul David-Bouldin (DB) este o măsură a calității clustering-ului în analiza clusterelor. Aceasta determină cât de bine sunt separate și distanțate grupurile formate de algoritmul de clustering într-un set de date.

Pentru a calcula indexul DB, se iau în considerare două componente principale: dispersia intra-cluster (intra-cluster dispersion) și dispersia inter-cluster (inter-cluster dispersion).

Dispersia intra-cluster măsoară cât de compacte și omogene sunt grupurile individuale. Se calculează prin medierea distanțelor dintre punctele dintr-un cluster și centrul său (de obicei, utilizând distanța euclidiană). Valori mici ale dispersiei intra-cluster indică grupuri strânse și omogene.

Dispersia inter-cluster măsoară cât de separat și diferit sunt grupurile între ele. Se calculează prin medierea distanțelor dintre centrul fiecărui cluster și centrul celorlalte clustere. Valori mari ale dispersiei inter-cluster indică grupuri distincte și bine separate.

Indexul DB se obține prin însumarea rapoartelor dintre dispersia intra-cluster și dispersia inter-cluster pentru fiecare cluster în parte, și apoi realizarea unei medii peste toate clusterele. Scopul este de a obține un index cât mai mic posibil, indicând o separare și distanțare mai bună între grupuri.

-indexul dunn

Indexul Dunn este o măsură utilizată în analiza clusterelor pentru a evalua calitatea unei soluții de clustering. Acesta ia în considerare atât compacitatea (compactness) cât și separația (separation) clusterelor.

Pentru a calcula indexul Dunn, se iau în considerare două componente principale: distanța minimă între două clusterelor diferite și cea mai mare dispersie internă a unui cluster.

Distanța minimă între două clusterelor diferite reprezintă cea mai mică distanță între oricare două puncte din clusterul diferit cel mai apropiat. Cu alte cuvinte, reprezintă distanța dintre cele două grupuri cele mai apropiate.

Dispersia internă a unui cluster reprezintă cea mai mare distanță între două puncte diferite din același cluster. Cu cât dispersia internă este mai mare, cu atât clusterul este mai puțin compact și mai dispersat.

Indexul Dunn se obține prin raportul dintre distanța minimă între două clusterelor diferite și cea mai mare dispersie internă a unui cluster. Scopul este de a maximiza acest index, ceea ce indică un clustering cu clusterelor bine separate și compacte.

-euristic=bazat pe experienta

-pt prostia de problema cu tabelul sigmoid(prag\*(prag \* suprafata locuinta + prag\* salar))

\*sigmoid are functie de activare

- hamming loss = 1/n \* suma(yPred(i) XOR yTrue(i)

-pt mutatie se pot folosi distributiile: Laplace, Cauchy

-sigmoid mapeaza orice nr real in intervalul (0,1)

A graph of a function

Description automatically generated with low confidence

-relu: in formă de rampă, continuă, monotonă, derivata ei este monotonă, codomeniu poziti

A picture containing line, plot, diagram, text

Description automatically generated

-regresia liniara e o tehnica de analiza statistica

**RETELE NEURONALE ATRIFICIALE**

-accepta atat date de intrare discrete, cat si continue

-au straturi de intrare, ascunse si de iesire

-functia de activare (relu) face 0 din valorile negative (le dezactiveaza)

-fiecare pixel e un neuron

-un neuron poate invata folosind regula perceptonului

-ce e un percepton?

-o retea neuronala isi imbunatateste ponderile in backpropagation

-entropia incrucisata e functie de loss

-un neuron calculeaza o functie liniara urmata de o functie de activare, ca daca ar fi invers degeaba ar calcula activarea

**DETECTARE COMUNITATI**

-ne trb locus-based ca asta e o reprezentare cu vecini care schimba informatii, child-based nu ne ajuta in cazul asta

-modularitate: Modularitatea în rețelele complexe se referă la cât de bine poate fi rețeaua împărțită în grupuri distincte de noduri. Este o măsură a structurii și organizării interne a rețelei. Rețelele cu modularitate ridicată au grupuri bine definite de noduri, indicând specializare și interacțiuni specifice în interiorul acestor grupuri. Modularitatea ajută la identificarea comunităților sau grupurilor în rețele și la înțelegerea interacțiunilor dintre noduri.