Genetisk Algoritme

Genetiske algoritmer er en metode, som er inspireret af Charles Darwins teori om naturlig selektion, som omhandler hvordan de biologiske arter udvikler sig over tid, ved at tilpasse sig miljøet og derved bliver bedre egnet til, at overleve og formere sig i miljøet. Teorien danner grundlag i, at populationen af en givende art har forskellige kromosomer og at der over tid vil ske små ændringer i kromosomerne hos individerne. Disse små ændringer i kromosomerne vil over længere tid, fører til større ændringer hos individerne. På daværende tidspunkt var Darwins teori meget kontroversiel, dette er dog ikke tilfældet for software brug af genetiske algoritmer, da en algoritme er noget lettere at forklare, end en biologisk ændring i en art over længere tid.

En genetisk algoritme består af følgende dele

1. Individer som er mulige løsninger til problemet.
2. Fitness som er en egenskab hos individerne, som bliver udregnet i forhold til, hvor god en løsning individerne er til problemet.
3. Et individ består af flere kromosomer, som er dele af et individ.
4. En population som består af en mængde af individer.
5. Generationer som indikere hvor lang tid algoritmer forløber over.

En population bliver muteret ved følgende operationer.

Selektion:  
Operatoren vælger individer til reproduktionen, jo større fitness individerne har, jo større sandsynlighed er der for, at de bliver valgt til reproduktion. Reproduktion kombiner 2 individer og danner ud fra deres kromosomer et nyt individ.

Crossover:

Operatoren vælger 2 tilfældige kromosomer og blander dem så der bliver dannet 2 nye kromosomer som er en kombination af de 2 kromosomer fx strengene 1110100 1011111 bliver krydset og danner de 2 nye strenge 1010100 1111111.

Mutation:

Operationen flipper tilfældigt nogle stykker af kromosomet fx strengen 1001000 bliver muteret i dens tredje position til 1011000. Mutation kan ske i alle positioner i strengen med en hvis sandsynlighed.

For en simpel algoritme kan fungere skal der være givet et defineret problem og et individ som en mulig løsning til problemet.

1. Der bliver tilfældet generet en population af n-individer med l-kromosomer
2. De genetiske algoritme operatorer udsætter populationen for mutation og derved vil der forekomme ændringer hos kromosomerne i individerne eller danne nye individer.
3. Der bliver beregnet fitness for hvert n-individ i populationen. Fitness bestemmer sandsynligheden for individet overlever i populationen.
4. Processen gentages fra trin 2 i x antal givende generationer eller indtil et givet kriterium er opfyldt.

Algoritmen er færdig når x antal generationer er kørt igennem eller et givet kriterium er opfyldt og der vil være en population med potentielle løsninger til problemet, hvor der herud fra vil blive valgt en løsning til problemet som normalt er løsningen med højst fitness.

Selektion

Roulette metoden

Man kan forestille sig at hvert individ er tildelt et stykke på en roulette og størrelsen på stykket er proportional med individets fitness. Roulette bliver spinnet n antal gange det vil tage for at vælge forældrene til den næste generation. Under hvert spin bliver individet under roulettens markør valgt til at være en del af en gruppe af forældre til den næste generation. En kandidat kan godt blive valgt til at være forældre flere gange, dette er okay da vi vælge forældre til næste generation og ikke selve individerne i generationen. Formålet med denne metode er at få valgt de forældre med den største fitness til næste generation da de har større sandsynlig for at skabe individer større fitness. Problemet med denne metode er dog at den genetiske algoritme hurtigt vil stå fast i den ene del af fitness rummet da det er muligt at vælge den samme forældre flere gange og derved kan der blive skabt en meget ens artet population som gør at der kun vil blive udforsket et bestemt område af rummet i stedet for at udforske hele rummet.

Rank metoden

Metoden har ligheder med roulette metoden med i stedet for at proportionel med den absolute fitness er den proportionel med den relative fitness. Der er altså ligegyldigt om den fitteste har 10 gange højere fitness en den næste i rangen eller om den har 0.0001% højere fitness. I begge cases vil sandsynligheden for den den fitteste være den samme.

Tournament metoden

2 tilfældige individer bliver valgt fra populationen. Man generer en tilfældig værdi fra 0-1 for sammenligner den med valgte sandsynligheds værdi. Hvis værdien er mindre eller lige med sandsynlighedsværdien bliver det individ med højst fitness valgt ellers bliver individet med den lavere fitness valgt. Sandsynlighedsværdien bliver altid sat højere end 0.5 for at favorisere individet med den højeste fitness.

”Seach space” referer til en gruppe af kandidat løsning til et problem, hvor der er en ”distance” i mellem kandidaterne. For eksempel lad os tage vigtigt problem indenfor bioengeering, hvordan man designer et protein. Antaget at man vil søge efter et protein som er en sekvens af aminosyre som kan blive brugt til at bekæmpe en virus. ”Seach space” vil være en kollektion af alle mulige proteiner. Dette vil give os uendelige mange muligheder derfor begrænser vi længden af proteinet til længden 50 som stadig vil være et stort ”seach space” siden der er 20 mulige aminosyre i hver position i proteinet. Hvis vi repræsenter aminosyrerne i form af alfabetet vil et muligt protein se ledes ud.

ASDKEGHB…. Vi definer distance mellem proteinerne som forskellen i alfabetet på den tilsvarende position i et andet protein fx ASDKEGHB og BSDKEGHB er distance 1 og distance mellem ASDKEGHB og GCCHAKAA er 8.

Fitness rum repræsentere alle mulige kandidater i rummet samtidig med deres fitness.