Corso di metodi computazionali, anno accademico 2020/21
Gruppo Python/algoritmi genetici

Tema d'esame

L'esame verterà sulla presentazione di un pacchetto software (in Python e utilizzando gli algoritmi genetici) preparato dallo studente, assieme a una breve (max 10 pagine) relazione. Per la presentazione, da svolgersi alla lavagna e/o con il supporto di slides, si dovrà prevedere una durata di circa 20 minuti. Software e relazione dovranno essere inviati al più tardi una settimana prima dell'esame.

Il software dovrà decriptare testi in lingua inglese (indicativamente della lunghezza di circa mezza pagina) codificati con diversi metodi di crittografia. Si prevedono quattro livelli, con difficoltà crescente. Affinché l'esame possa considerarsi passato, occorre risolvere almeno il primo livello.

1. crittografia poli-alfabetica con chiave di lunghezza nota (tra i 4 e i 15 caratteri)
    - Fissato un numero massimo di generazioni, come scala il numero di individui necessario ad arrivare a una soluzione?
    - Fissati numero di generazioni, individui e lunghezza chiave, per che range di L(testo)/L(chiave) la decodifica funziona?
2. crittografia mono-alfabetica (l'alfabeto di cifratura è una permutazione dell'alfabeto originale)
    - Quando il testo diventa comprensibile anche se non completamente decrittato?
3. (può essere svolto subito dopo il punto 1) crittografia poli-alfabetica con chiave di lunghezza ignota (sempre compresa tra i 4 e i 15 caratteri)
4. crittografia non nota, di tipologia compresa tra quelle sopracitate.

In tutti i casi l'alfabeto può includere o meno il carattere spazio. Nel secondo caso, gli spazi saranno rimossi dal testo in chiaro prima della codifica. Non si dovrà fare differenza tra lettere maiuscole e minuscole.
Il file contenente il testo codificato sarà passato come primo argomento all'eseguibile del pacchetto. La possibilità di passare ulteriori argomenti è ammessa e dovrà essere opportunamente documentata.

Suggerimenti:

- Gli individui dell'algoritmo genetico saranno le chiavi di cifratura (permutazione dell'alfabeto nel caso mono alfabetico o parola chiave nel caso poli alfabetico)
- Per quanto riguarda la funzione di fitness, essa può essere basata sul conteggio delle sillabe più frequenti in lingua inglese (combinazioni di due o tre lettere) e eventualmente delle parole brevi nel testo decrittato. Si veda per esempio https://www3.nd.edu/~busiforc/handouts/cryptography/cryptography%20hints.html
- Eventualmente si possono prevedere penalità alla funzione di fitness se ci sono (molte) parole di 2 o 3 lettere che non sono tra le più frequenti
- Per risolvere il punto 2, conviene che lo spazio sia nell'alfabeto. Inoltre, si può scegliere la popolazione iniziale non completamente random, ma che ad esempio non dia parole (sequenze di caratteri comprese tra due spazi) troppo lunghe nel testo decrittato, e/o abbia un fitness maggiore di una certa soglia
- È bene, nei run di prova del software, monitorare che il fitness del testo in chiaro sia sempre superiore a quello degli individui della popolazione, e che ci sia correlazione tra fitness e numero di posizioni indovinate all interno della chiave

- Per quanto riguarda le richieste ai punti 3 e 4, una possibile strategia può essere quella di creare AG con sotto-popolazioni, per esempio una sotto-popolazione per ciascun valore della lunghezza della stringa. A ogni generazione, il numero di individui di ciascuna sotto-popolazione dipenderà dal fitness globale della medesima.

In caso il software non fosse funzionante, è comunque ammesso presentarlo purché si discutano in dettaglio le problematiche e eventuali strategie per risolverle.