**תרגיל מסכם**

**חלק א': בחירת אלגוריתמים ותיאורו**

מאמר ראשון :   
M. Mohammadi, H.S. Noghabi, G.A. Hodtani, H.R. Mashhadi Robust and stable gene selection via maximum–minimum correntropy criterion Genomics, 107 (2–3) (2016), pp. 83-87

המאמר מתאר דרך לבחירת פיצ׳רים על ידי פתירת בעיית אופטימיזציה, כאשר בפונקציית המטרה ישנם שני מקדמים ששינוי שלהם יכול להביא לבחירת פיצ׳רים מוצלחת יותר.

לפי הטקסונומיה המתוארת בתרשים 3 במאמר Hambali et al. [2020] המאמר שייך לאלגוריתמי חיפוש מסוג פילטר רבת משתנים ((multivariate.

צעדים:

1. חישוב Correntropy עבור כל פיצ׳ר מול ה class על ידי שימוש במקדם σ
2. חישוב Correntropy עבור כל פיצ׳ר מול כל שאר הפיצ׳רים והכפלה של הסכום במקדם β
3. חשיבות כל פיצ׳ר מחושבת על ידי ההפרש בין התוצאה מצעד 1 לבין התוצאה מצעד 2
4. נבחרים K הפיצ׳רים החשובים ביותר
5. מתבצע עבורם הערכת דיוק על ידי SVM
6. משנים את σ ואת β וחוזרים לצעד 1
7. לאחר X חזרות מחזירים את הפיצ׳רים שנתנו את ההערכה הטובה ביותר בצעד 4

היתרונות

1. מתחשב בכל הפיצ׳רים, ישנם אלגוריתמים שלא בודקים את כל הפיצ׳רים בגלל שהם עובדים מגרילים באיזה פיצ׳רים להשתמש, באלגוריתם הזה מחשבים את ה Correntropy עבור כל הפיצ׳רים.
2. ה Correntropy הוכח כמדד טוב לחיזוי טיב הפיצ׳ר
3. פשוט וקל להבנה ומימוש

חסרונות

1. ה Correntropy הינו מדד יוריסטי אשר במקרים מסוימים יכול להטעות והפיצ׳ר כלל אינו רלוונטי למרות שקיבל ציון גבוהה במדד.
2. זמן ריצה איטי מאוד עבור כמות גדולה של פיצ׳רים.

**הרצה לדוגמא SPECTF.train**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| iteration | sigma | alpha | top 10 features | | | | | | | | | | score |
| 1 | 1880 | 0.024 | 44 | 43 | 26 | 25 | 28 | 16 | 27 | 33 | 42 | 10 | 0.75 |
| 2 | 2853 | 0.019 | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 0.81 |
| 3 | 1385 | 0.023 | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 0.76 |

ניתן לראות כי הפרמטרים שנתנו את הערך הגבוה ביותר הן של איטרציה 2

**חלק ג' – שיפור**

השיפור המוצע הינו על ידי שיוני בפונקציית חיפוש המקדמים, ההצעה שלי הינה לשנות מחיפוש differential evolution לחיפוש מבוסס gradient optimization.

השיפור הביא בעיקר לשינוי בזמן ריצה של האלגוריתם – זמן הריצה התקצר בין 50% - 70%, שיטת החיפוש הקודמת הינה שיטה סטוכטית שאינה לוקחת בחשבון שיש משמעות לטווח ערכים מסוים עבור כל אחד מהמקדמים.