היי שאול,

מימשתי מערכת לומדת, המחשבת את המודל הלומד הטוב ביותר כולל ההיפר פרמטרים הכי טובים, עבור הדאטה: German Credit שלקחתי מ Kaggle.

המודלים הנבדקים (כולל ההיפר פרמטרים הייחודיים שלהם) הם:

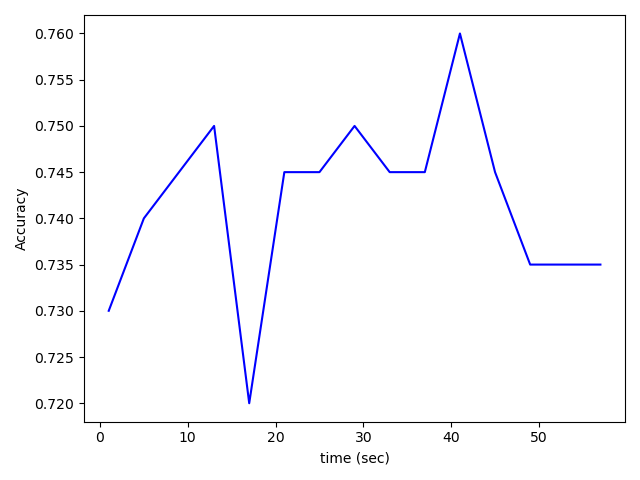
1. Random forest
2. SVM
3. KNN
4. Ada Boost

מצורף גרף ראשוני של דיוק מודל נבחר (test accuracy) עבור הזמן המושקע בבחירת התכונות (feature selection).

בחירת התכונות ע"י האלגוריתם מתבצעת כך:

1. ראשית יבוצע אלגוריתם חיפוש מקומי קדמי.
2. אם נשאר זמן, יבוצע אלגוריתם חיפוש מקומי אחורי
3. אם נשאר זמן, האלגוריתם מתחיל לבדוק את כל הקומבינציות האפשריות של תכונות, עד שנגמר לו הזמן.

לבסוף, האלגוריתם בחר את התכונות שנתנו את הדיוק הטוב ביותר. הדיוק נבדק על חלק סטרילי של ה training set ומבוצע ע"י אלגוריתם KNN.



ניתן לראות כי עבור 40 שניות של חיפוש תכונות טובות ביותר, מקבלים דיוק טוב ביותר על ה test set. כלומר, מה שאני מבין, שקיים זמן מסויים של חיפוש תכונות, שאחריו