נשים לב כי אנו מנסים לפתור בעיית soft svm תחת התנאים הבאים:

* ע״פ הגרף המצורף לשאלה, הדאטה לא נראה פריד לינארית ולא נראה כי ניתן להשיג מידת דיוק גבוהה על הדאטה עם מפריד לינארי.
* ההיפר פרמטרים המצורפים לשאלה הם C שהוא יחסית גדול מאד, וlearning rate יחסית קטן וקבוע.

ניתן לראות כי במהלך תהליך האימון, ע״פ התרשים, ערך פונקציית הloss הלך וירד ,כאשר בהתחלה הייתה ירידה חדה ולאחר מכן הירידה איטית יותר ומתכנסת לערך מסויים.

ציפינו שירידה זו תקרה במהלך האימון כיוון שלבעיית הsoft-svm קיים מינימום גלובאלי כפי שראינו בהרצאה ובתרגולים, ולכן ע״י שימוש באלגוריתם sgd ערך הפונקציה צפוי לנוע לכיוון הערך המינימלי במהלך האימון (עם זאת ,זה לא מובטח כי פונק׳ המטרה לא גזירה וlr קבוע). ה learning rate שאנו משתמשים בו יחסית קטן, ולכן הירידה איטית יחסית וגם אין התבדרות.

מאחר שC גדול מאוד, ע״פ הגדרת בעיית האופטימיזציה, תינתן עדיפות לצמצום פונק׳ הhinge על פני מציאת מפריד עם margin גדול. עם זאת, מאחר והדאטה אינו פריד לינארית, ניתן לראות כי ערך פונ׳ הloss נשאר יחסית גבוה בסוף האימון, כלומר ערך פונקציית הhinge לא הצליח להגיע לערך מאד נמוך ולכן גם הloss נשאר גבוה.

מבחינת הaccuracy, ניתן לראות כי במהלך האימון היו עליות וירידות ולא הייתה התכנסות למידת דיוק כלשהי. נזכיר כי הaccuracy הוא המשלים של ה0-1 loss .

כאשר אנו פותרים בעיית soft-svm אנו כן מקווים שפתרון לבעיית svm יצליח להביא גם לשיפור במידת הדיוק. עם זאת, פונקציית המטרה בבעיית הsoft-svm מכילה את פונקציית הhinge ולא את פונקציית ה0-1 loss. לכן, פתרון של בעיית soft-svm אינו מבטיח פתרון אופטימלי עבור 0-1 loss.

עוד נזכיר כי ראינו בהרצאה, שפונק׳ הhinge היא קירוב של פונק׳ ה0-1 loss , אך כאשר אנו מסתכלים על אזורים בהם הנקודות כבר בצד הלא נכון, הפונקציות האלה כבר לא מאוד קרובות כיוון שפונק׳ הhinge מענישה את הנקודות שלא מתויגות נכון בעונש כבד יותר ככל שהן רחוקות יותר מהmargin.

כאשר אנו מחפשים את נק׳ המינימום של בעיית הsoft-svm ייתכן מצב שבו כדי להקטין את הloss , עדיף פשוט לנסות למצוא מפריד שיותר קרוב גם לנקודות שאינן מתויגות נכון מאשר לנסות להגדיל את כמות הנקודות באימון שמתויגות נכון.

לכן, במצב של הדאטה שלנו, שהוא אינו פריד, ויש הרבה דגימות שיכולות להיות רחוקות מכל מפריד אופציונאלי, פונק׳ הhinge היא למעשה קירוב כבר לא כ״כ טוב ל0-1 loss . לכן ,תופעת התנודות של הaccuracy שראינו בתרשים , היא הגיונית, תחת התנאים שהזכרנו במהלך הדיון.