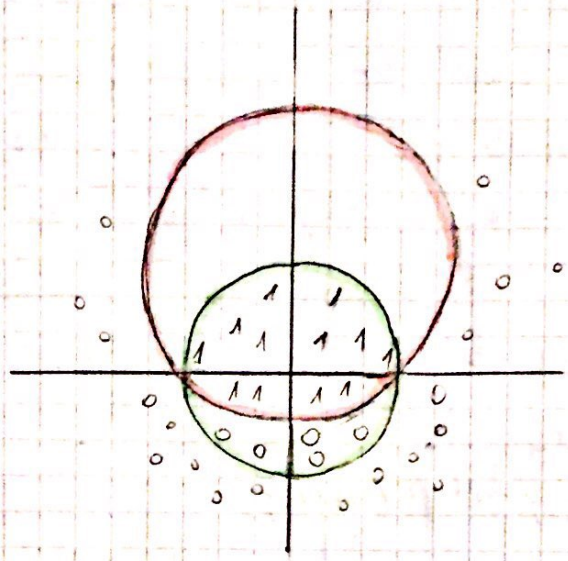
אלמוג לק
307900183

שאלת מחשבה - תרגיל 1

1.1 אלגוריתם A מחזיר ERM. נשתמש ב-A כדי להלן הבאה:



- עבור ה- training set המוצג בשרטוט (ה- 1, 0) האלגוריתם A יחזיר את המסווג הירוק. שכן הוא מגיע הכי הרבה הטעויות שמכיר את הנקודות המסווגות כ- 1. ויחד עם זאת הוא כולל הכי הרבה טעויות מכל המסווגות כ- 0, ולכן $loss_S = 0$.
- לעומת זאת- המסווג הירוק הוא מסווג בעל הכי הרבה טעויות מכל הירוק, ומכיל רק את הנקודות המסווגות כ- 1 ולכן $loss_S = 0$ - כלומר אין שגיאה. והוא מסווג בעל הכי הרבה טעויות מכל הירוק (שהוא מסווג אחר עם $loss_S = 0$).

$\Rightarrow$ האלגוריתם A אינו אלגוריתם ERM* מכיוון שקיימים מסווגים (כמו זה שבשרטוט שמעלה) שעבור מסווג אחרים ש- A החזיר נקבל $loss_S$ שהוא הכי הטעויות - כלומר השימוש ב-A נקבל שמסווג אחרים עבר מבחינת מבחר של ה- $loss_S$ לטעויות בירוק ולכן A הינו לא ERM.

* אלגו ERM הכי שמחפש את אחד הסווגים שמסווגים את השגיאה (ה- $loss_S$). (במידה שיש A מחזיר מסווג אחרים שגם ה- $loss_S$ לא טעויות ולכן A גם ERM).

הערה: הדרה האקסיומטית עבורה כן נקבל ERM היא **מחלקה** בעלת מכיוון שהשיטה מסווגים ניתן לשנות- שאנו מעוניינים להשפיע על הרווח

כנ"ל ניתן לשלול חלק מהפתרונות בקלות. כמו כן אלגוריתם שמחזיר את הפתרון שנמצא בקבוצת ההיפותזות שלו ובמקרה שבו אין פתרון מחזיר (למשל) 0, נשאר כמובן בעיה שבה ניתן לשנות את הפתרון ולקבל שגיאה גדולה (למשל), ולכן יש לנו יותר מדי פתרונות שבהם הערך האמיתי של $L_D(h)$ הוא 1 בהתאם לדוגמה הקודמת. כיוון שאין למשל דרך לדעת אם ציר הפתרון הוא מוסף או לא, נדרש עבור אלגוריתם כזה לקבל את "מיטב" הפתרונות כלומר כאלו שהשגיאה האמיתית שלהם היא 0. אפשר גם לחשוב על "הדוק" כך ניתן לשנות את האלגוריתם כך שיחזיר פתרון שהשגיאה האמיתית שלו היא 1 (וזה בעצם מה שעושה אלגוריתם ERM שאנו נתקלים בו בהמשך).

1.2

$$\mathbb{E}_{S\sim D^m}[L_S(h)] = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}_{[h(x)\neq f(x)]}\right] =$$

$$= \mathbb{E}\left[\frac{1}{m}\cdot m\cdot\mathbb{1}_{[h(x)\neq f(x)]}\right] = \frac{1}{m}\cdot\mathbb{E}\left[\sum_{i=1}^{m}\mathbb{1}_{[h(x)\neq f(x)]}\right] = \frac{1}{m}\cdot m\cdot\mathbb{E}\left[\mathbb{1}_{[h(x)\neq f(x)]}\right] =$$

(לינאריות התוחלת) (לינאריות התוחלת)

$$= \mathbb{E}\left[\mathbb{1}_{[h(x)\neq f(x)]}\right] = 1\cdot P_{x\sim D}[h(x)\neq f(x)] + \underbrace{0\cdot P_{x\sim D}[h(x)=f(x)]}_{0} =$$

(תוחלת של משתנה אינדיקטור)

$$= P_{x\sim D}[h(x)\neq f(x)] = L_D(h) \quad \blacksquare$$

Scanned with CamScanner