**תרגיל בית 3 רטוב**

**מבוא למערכות לומדות**

**תאריך הגשה – 28.06.2022**

מגישים:

יואב יעבץ, תעודת זהות 212617864.

מור לוי, תעודת זהות 211810452.

**חלק 1**

Chart, line chart, histogram

Description automatically generatedChart, line chart

Description automatically generatedשאלה 1:

שאלה 2:שני הגרפים המצורפים למעלה מתארים את חלוקת ערך מאפיין המטרה CovidScore בהתאם למאפיין BloodType, כאשר הגרף מימין מבצע חלוקה זו לפי כל סוג אפשרי של המאפיין BloodType והגרף משמאל עושה זאת לפי חלוקה לכמה קבוצות של סוגי המאפיינים.

נשים לב כי הגרף מצד ימין מראה לנו כמה תכונות אשר מתקיימות:

* גרף הצפיפות של מאפיין המטרה CovidScore של מטופלים עם דם A+ או עם דם A- דומה אחד לשני.
* גרף הצפיפות של מאפיין המטרה CovidScore של מטופלים עם דם O+ או עם דם O- דומה אחד לשני.
* גרף הצפיפות של מאפיין המטרה CovidScore של מטופלים עם דם B+ או עם דם B- או עם דם AB+ או עם דם AB- דומה אחד לשני.

לכן הגיוני להוסיף לחלק את סוג הדם לשלוש קבוצות, כפי שתואר בשאלה 1, כי בכל קבוצה גרף הצפיפות המתקבל של מאפיין המטרה CovidScore דומה לגרף הצפיפות המתקבל של מאפיין מטרה זה ללא איחוד סוגי דם אילו לקבוצה אחת. משמע לא איבדנו מידע באיחוד קבוצות זה והורדנו את מספר המאפיינים – מה שמוריד את סיבוכיות מחלקת המודולים שלנו, ויכול להביא לתוצאות יותר טובות בעתיד.

**חלק 2**

שאלה 3: מתקיים . כעת:

Chart, line chart

Description automatically generatedשאלה 4:

שאלה 5: גרף מצורף למטה. ניתן לראות כי מתקיים:

* עבור ערכי learning rate קטנים מ- מתקבל כי ה-MSE Loss על ה-train data ו -Validation data קבוע במקומו (עד כדי ירידה זניחה). התנהגות זו הגיונית, כי ניזכר שנוסחת העדכון של SGD היא , לכן אם ה-learning rate קטן מאוד לא נעדכן את הפרמטרים, לכן ה-MSE Loss לא ישתנה.
* עבור ערכי learning rate גדולים מ- וקטנים מ- ניתן לראות כי ה-MSE Loss על ה-train וה-validation כתלות במספר ה-iteration יורד באופן עקבי. זו התנהגות הגיונית כי אנחנו מצפים שעבור ערכי למידה טובים אלגוריתם SGD יעדכן את סט הפרמטרים כך שה-MSE Loss ירד עם הזמן.  
  נוסף על כך, עם גדילת ערך ה-learning rate לכיוון 0.01 ניתן לראות כי ערך ה-MSE Loss עבור ה-train וה- validation אליו מגיעים לאחר 1500 איטרציות נמוך יותר. זו גם כן התנהגות הגיונית, הרי ברור כי ערכי למידה שונים יגררו קצב למידה שונה. עבור כל ערכי הלמידה ניתן לראות ירידה ב-MSE Loss אבל קצב הירידה שונה לכל אחד, כי עדכון סט הפרמטרים באלגוריתם שונה לכל אחד. יכול להיות כי עבור כל ערכי למידה אלו עבור מספר איטרציות שואף לאינסוף נתכנס לאותו ה-loss, אך כל עוד זה לא מתקיים – קצב הלמידה שונה ולכן גם ה-loss אליו מגיעים לאחר 1500 איטרציות שונה.
* עבור ערך learning rate השווה ל-0.1 ניתן לראות כי ה-MSE Loss על ה-train וה-validation יורד עם הזמן, אך אינו יורד בצורה מונוטונית - קיימות קפיצות בערך ה-loss עבור שני הסטים של המידע באיטרציות שונות לאורך למידת האלגוריתם. דבר זה קורה כתוצאה מבחירת learning rate גבוה מדי, מה שמוביל לעדכון סט הפרמטרים בצורה גדולה מדי כאשר מתקרבים אל נקודת המינימום בתחתית. לכן אנחנו לא מצליחים להתכנס אל סט הפרמטרים שיביא את ה-loss נקודת המינימום, כי מדלגים מעליה כל פעם (הערה - כל פעם הדילוג מקרב אותה אל נקודת המינימום של ה-loss עוד טיפה).

Graphical user interface

Description automatically generatedקצב הלמידה הטוב ביותר הוא בין 0.01 ל-0.1 לפי הגרף שנראה מטה.  
חשוב לציין כי יש היגיון להגדיל את מספר ה-gradient steps עבור קצב הלמידה הטוב ביותר.   
עבור קצב זה אנחנו רואים התכנסות אל עבר סט הפרמטרים שמוביל אותנו לנקודת המינימום של ה-loss, והתכנסות זו היא באופן מונוטוני ללא קפיצות. לכן הגדלת מספר איטרציות העדכון יוביל אותנו לעדכון וכוונון סט הפרמטרים כך שנגיע לטוב ביותר המביא לנו את ה-loss המינימלי.

**חלק 3**

שאלה 6: