דף עבודה - חישוב פונקציית XOR באמצעות רשת נוירונים רדודה

דף עבודה מוגש על ידי: בתאריך:

את דף העבודה יש למלא בעזרת הרצה של הקוד הנמצא בקבצים הבאים:

XorUsesNN.py, NN.py

שימו לב: בכל מקרה בו אתם נדרשים לרשום ערכי משתנים, יש לרשום מה המשמעות של כל משתנה, ולמעט נגזרות יש גם לרשום את החישוב שהביא לערך זה ואת הערכים אותם יש למלא בדיוק של לא יותר מ- 3 ספרות אחרי הנקודה.

1. כמו בכל בעייה בה נתקלנו עד עכשיו – אנחנו צריכים לבחור את המודל, לאפיין את הקלט והפלט, לאמן את המודל כלומר לחשב את ערכי W,b ואחר כך להשתמש במודל ובמשקלים שחישבנו על מנת לבדוק את המודל ולבצע חיזוי.
2. על מנת להראות את כל הפרטים (הרבים!) הקיימים ברשת נוירונים ולו הקטנה ביותר, נתחיל עם בעייה פשוטה יחסית – בעיית ה-XOR. ראינו שבעייה זו לא ניתנת לפתרון על ידי "נוירון" (נקרא גם פרספטרון ובעצם זו רגרסיה לוגיסטית) אחד.

1. מהם ערכי הקלט לבעיית XOR?

נזכור כי עבור 2 משתנים בוליאנים יש לנו 4 קלטים אפשריים כי כל אחד מהמשתנים יכול לקבל 2 ערכים, אמת או שקר (כלומר 0 או 1).  כל אחד מהקלטים מוצג כעמודה ויש לו את הפלט המתאים לו.

רישמו כאן את הערכים של עמודות הקלט השונות:

ואילו כאן רישמו את הפלט המתאים לכל אחת מהעמודות:

1. כאמור הבחירה במודל (רגרסיה לינארית, רגרסיה לוגיסטית, רשת נוירונים וקיימים עוד מודלים רבים אחרים בבינה מלאכותית) הינה החלטה שקיבלנו מראש. רק כשנחשב את תוצאות המודל עבור קלט שאנחנו יודעים מראש את הפלט שלו ונשווה את הפלט שקיבלנו מהמודל לפלט הידוע לנו (כלומר פרוצדורת Test) נוכל לדעת אם המודל מצליח לפתור את הבעייה במידת הדיוק הדרושה לנו.

במקרה שלנו הבחירה היא ברשת נוירונים (לאחר שראינו שרגרסיה לוגיסטית לא מצליחה למצוא משקלים מתאימים עבור XOR – ואחרי שהבנו גם את הסיבה. רישמו כאן את הסיבה מדוע רגרסיה לוגיסטית, כלומר נוירון אחד, לא יכול לפתור את בעיית ה-XOR:

1. רשת הנוירונים שלנו נראית כך:

Diagram

Description automatically generated

מה גודל הקלט בכל דוגמא (כלומר מה גובהה של עמודה אחת בקלט בתוכנה עצמה)?

כמה נוירונים בשכבה הסמויה, Hidden Layer?

כמה נויורונים בשכבת הפלט?

שימו לב שההחלטות לגבי מבנה הרשת, כמות השכבות וגודלן, הן בחירה שרירותית שלנו. ככל שנעמיק ונצבור ניסיון נדע טוב יותר להתאים את הרשת לבעייה אותה אנו רוצים לפתור.

1. מלבד בניית הרשת, האתחול כולל גם קביעת משקלים, במקרה שלנו על ידי בחירת משקלים רנדומוליים מהתפלגות אחידה. רשמו כאן את הערכים בהם בחרה התוכנה בבחירה הרנדומלית (לאחר קביעת seed) **עבור השכבה הראשונה:**

W11= W21=

W21= W22=

מהם ערכי b הראשונים של השכבה הראשונה?

b11= b12=

ואילו עבור **השכבה השנייה,** כמה ערכי w יש לנו? כמה ערכי b? איך הם מסומנים ומה הערכים ההתחלתיים שלהם?

1. בדומה למודלים שראינו עד עכשיו בקורס, רשת הנוירונים משתמשת ב-Gradient Descent על מנת למצוא משקלים המביאים למינימום את פונקציית ה-cost בה בחרנו. במקרה של רשת אנחנו מבצעים את החישוב המלא של הרשת עובר כל קלט ("חלחול קדימה" Forward Propagations) ואז מתבצעת השוואה לפלט הנתון מראש וישנו "חלחול לאחור" Backward Propagation) כלומר חישוב הנגזרות של המשקלים ועדכון של הערכים האחרונים בהם השמשנו. בדיוק כמו שעשינו עד עכשיו גם במקרה זה נחזור על כל "סיבוב" כזה (חלחול קדימה, חלחול אחורה, עדכון) הרבה מאוד פעמים עד שלא יהיה יותר שינוי משמעותי בערך פונקציית ה-cost או כמות שרירותית וגדולה ושל סיבובים.
2. נרשום כאן את ערכי הסיבוב הראשון עבור הקלט 0, 0:

בשכבה הראשונה, עבור הקלט 0,0 מהם הערכים הבאים? יש לרשום את המשוואה בה השתמש הקוד ואת הערך שהתקבל בדיוק של 3 ספרות אחרי הנקודה:

z11= a11=

z12= a12=

ואילו עבור השכה השנייה (יש למלא באותו האופן כמו בשכבה הראשונה):

z21= a21=

1. כעת נעקוב אחרי החלחול לאחור. מה הערך שהרשת מצאה עבור הקלט 0,0 בסיבוב הראשון בו היא השתמשה מנשקלים אקראיים? מה הערך הנכון לקלט זה?
2. אילו ערכים התקבלו עבור הקלט 0, 0 בחלחול לאחור הראשון עבור המשתנים הבאים? מה המשמעות של כל משתנה כזה?

dz21= da21=

dz12= da12=

dz11= da11=

1. מה ערכי הנגזרות שהתקבלו בסיבוב הראשון עבור הקלט 0,0?

dW21= dW22= db21=

dW11= dW12= db11=

dW21= dW22= db12=

1. מהם המשקלים החדשים לאחר העדכון הראשון? העזרו בשאלה 6 בדף עבודה זה (משקלים במצב הפתיחה) כדי לדעת לאיזה משתנים הכוונה.

1. ראינו בפירוט רב כיצד מתבצע ראשון של עדכון. כעת נעבור לסיום התהליך. איזה משקלים מצא המודל בסיום פרוצדורת fit?

W11= W21= b11=

W21= W22= b12=

W21= W22= b21=

1. כעת נבדוק האם המודל (רשת נוירונים) והמשקלים שמצאנו אכן מחשבים נכון את פונקציית XOR. כיון שיש לנו רק 4 קלטים שונים, נעביר לפרוצדורת test כל אחד מ-4 הקלטים השונים ונבחן את התוצאה שהמודל חישב לכל אחד מהקלטים.כלומר בניגוד למה שבדרך כלל עשינו ונעשה – לא נבחן את המודל על ערכי test חדשים, פשוט בגלל שבמקרה של XOR אין לנו עוד קלטים. רק במקרה זה ולמטרות לימוד אנחנו עושים מה שנקרא test on train (שזה בדרך כלל שגיאה חמורה). רישמו כאן את כל החישובים שהתבצעו עבור הקלט 0,0:
2. האם המודל שבנינו, רשת הנוירונים הרדודה, מצליח לחשב נכון את פונקצית XOR?
3. כיון שזו הפעם הראשונה בה אנו משתשמים בדף עבודה זה- אני בקרן לגבי תחושת ההבנה שלכם. האם אתם מרגישים שהבנתם כיצד פועל הקוד? האם אתם חושבות וחושבים שזו דרך לימוד אפקטיבית? בהזדמנות זו אתם מוזמנים לכתוב כל פידבק שנראה לכם רלוונטי.

דף העבודה הוכן על ידי עמיר שחל, מרץ 2022.