**ממן 12+13**

**נושא הממן: סיווג ספרות**

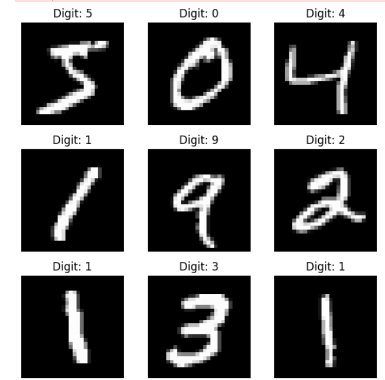
**מטרה:** מטרת הממן היא לאמן את המחשב בזיהוי ספרות אשר נכתבו בכתב יד

**השיטה:** בממן 12 נשתמש ב Softmax regression. נאמן את המודל שלנו ע"י חישוב ההסתברויות השונות לכל מחלקה עד אשר יתכנס ונחזה את התוצאה הסבירה ביותר  
בממן 13 נשתמש ברגרסיה לינארית נקבל מטריצת משקולות ע"י פתרון המשוואות הנורמליות ונחזה לפי התוצאה הטובה ביותר

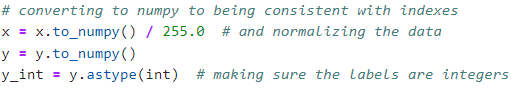
**אלגוריתמים מרכזיים:** שימוש ב Softmax regression ו linear regression

**ספריות שבהם השתמשתי:**

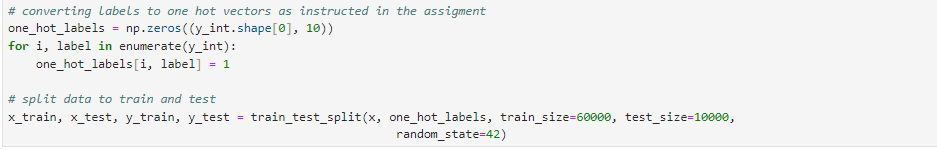
* Numpy – ספריה מתמטית אשר מיועדת לחישובים על מטריצות וקטורים וכו'...
* Matplotlib – ספריה שנועדה לסרטוט של הגרפים השונים
* Sklearn – ספריה שממנה מגיעים הנתונים ומאפשר לבצע מניפולציה על הנתונים.

**הנתונים:** תמונות של מספרים בגודל 28X28 פיקסלים ב gray scale ולכל אחד תווית אשר אומרת איזה מספר מיוצג בתמונה:

**הכנת המידע:** הכנת מידע בוצעה ע"י העברה של כל המידע למטריצות numpy (מגיע בתור pandas series) ונרמול של הפיקסלים לערכים של 0-1 במקום 0-255



לאחר מכן התוויות הומרו לתצורה של one hot



תמונה שמכילה טקסט, גופן, צילום מסך

התיאור נוצר באופן אוטומטי

המידע חולק ל 60 אלף תמונות אימון ו 10 אלף תמונות מבחן כפי שהתבקש בממן. בנוסף היה צורך להוסיף את איבר ההטיה.

**האלגוריתם:**

ממן 12:

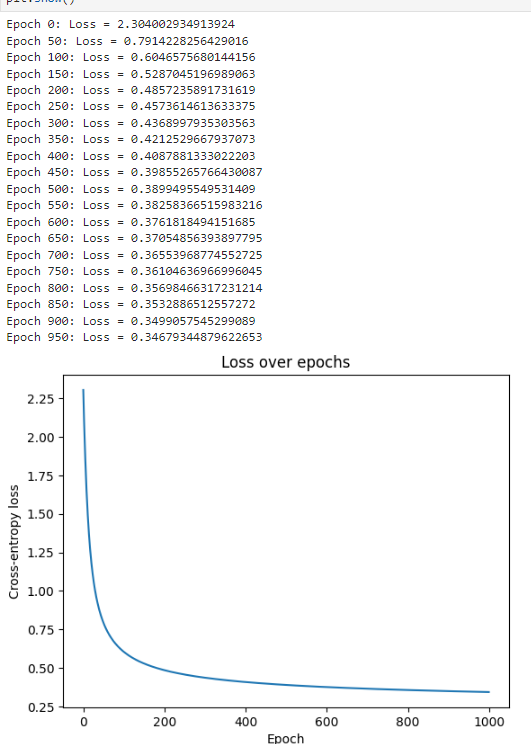
האלגוריתם שלנו עובד בצורה שהוא מנסה למזער את ה loses עד לכדי התכנסות   
בכל ריצה נקרא לפונקציית ה softmax ונתקן את האלגוריתם שלנו לכיוון של התוצאה על פי ה softmax שקיבלנו לעומת התוצאות האמתיות

ממן 13:

ע"י פתרון המשוואות הנורמליות נקבל ס"ט משקולות אשר יהווה לנו חזאי.

**ריצת התוכנית:**

עבור ממן 12 אנו רואים כי לאחרי 100 איטרציות ה loss שלנו כבר מתחיל להתכנס לערך ולאחר 100 איטרציות יש לנו כבר התכנסות לכיוון הערך 0.346



ובאופן כללי האלגוריתם שלנו מציג אחוזי ניבוי מרשימים של כ 90%



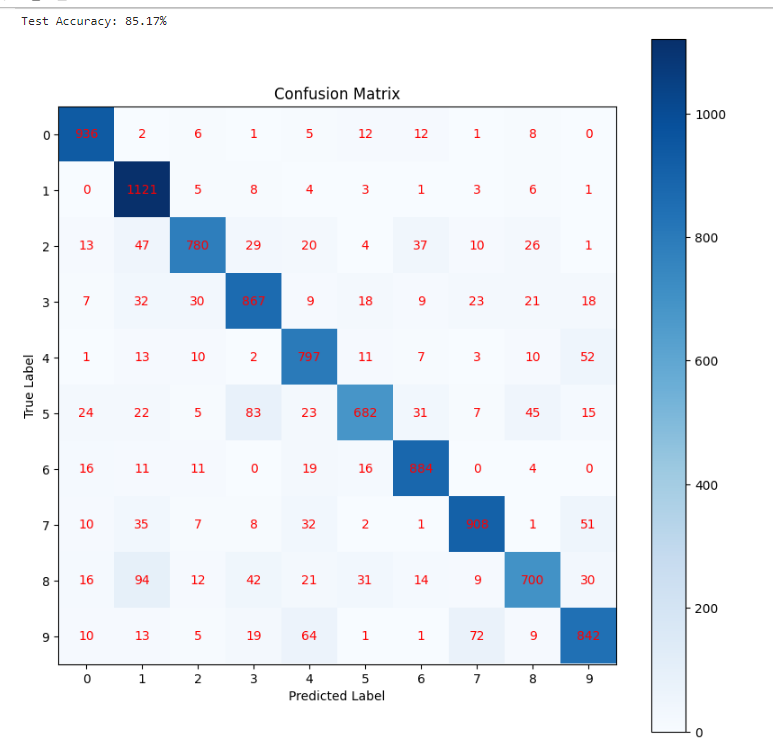
תמונה שמכילה טקסט, צילום מסך, תרשים, ריבוע

התיאור נוצר באופן אוטומטי

עפ"י ה confusion matrix ניתן לראות כי האלגוריתם שלנו מנבא סה"כ באופן דיי טוב את הספרה 0 נראה כי רוב הטעויות נופלות בין 3 ו 5 טעות זו הגיונית כי 3 ו 5 ספרות שהחלק התחתון שלהם דיי וזה והחלק העליון של 3 הוא מעוגל ו ב 5 הוא ישר מה שיכול לגרום לטעויות בזיהוי. ובין 8 ל 3 מסיבה דומה.

ממן 13:

פתרון המשוואות הנורמליות יניב לנו סט משקולות אשר הפעלתו על נתוני המבחן יניב לנו תוצאות של כ 85%



עם טעויות דומות לממ"ן אך נראה כי יש גם הרבה טעויות בין 4 ו 9 מסיבה שגם הן ספרות שדיי דומות

לסיכום

לסיכום. ניתן לראות כי שימוש ב softmax ניבא לנו את התוצאות הטובות ביותר אך אלגוריתם זה יכול להיות מסובך יותר וכבד יותר מאשר pocket algorithm שמביא תוצאות כמעט זהות לחלוטין

לעומת זאת אנו יכולים לראות כי אלגוריתם הרגרסיה הלינארית אינו טוב במיוחד למשימה הזאת ודיי מוגבל מכיוון שאני מסתמכים על פתרון משוואות נורמליות שהיו תלויים מאוד בנתונים שלנו ולא בונים מודל שמנסה לשפר את עצמו מריצה לריצה כמו ב 2 האלגוריתמים האחרים ולכן קיבלנו אחוזי ניבוי קטנים יותר עם זאת הוא הרבה יותר פשוט למימוש.