

3.א. הרשת שנבחרה מבוססת על פונקצית מחיר מסוג MSE. התוצאות המיטביות התקבלו בעזרת שכבת קלט בעלת 2 נירונים, 2 שכבות "נסתרות" בעלות 30 נירונים כל אחת (31 כולל נירון בייאס) ושכבת פלט בעלת 2 נירונים. לבסוף בחרתי להשתמש בפונקציות אקטיבציה מסוג Tanh לכלל השכבות מלבד האחרונה ובאחרונה להשתמש בsigmoid. שימוש בהגרלת משקלות באמצעות Xavier שיפר משמעותית את ביצועי הרשת ולכן נעשה שימוש בו.

ב. כאשר הוספתי איבר מומנטום לכלל העדכון, ניכר כי קצב ההתכנסות היה מהיר ויעיל יותר. ניתן לבחון את ההשפעה באמצעות הזנה $\alpha = 0$.

ג. בהוספת רגולריזציה ניכר כי שגיאת ההכללה פוחתת. בתצורה הנ"ל, קיימים 3600 קשרים. במצבים בהם יש כמות קשרים גבוהה ממספר דוגמאות האימון (1400) מומלץ לפשט את המודל על ידי חיזוק הקשרים החזקים והחלשת החלשים וזאת באמצעות רגולריזציה. תהליך זה אמנם עלול לפגוע בשגיאת האימון, אך בד"כ הוא תורם ליצירת שגיאת הכללה נמוכה יותר (ולתוצאות גבוהות יותר).