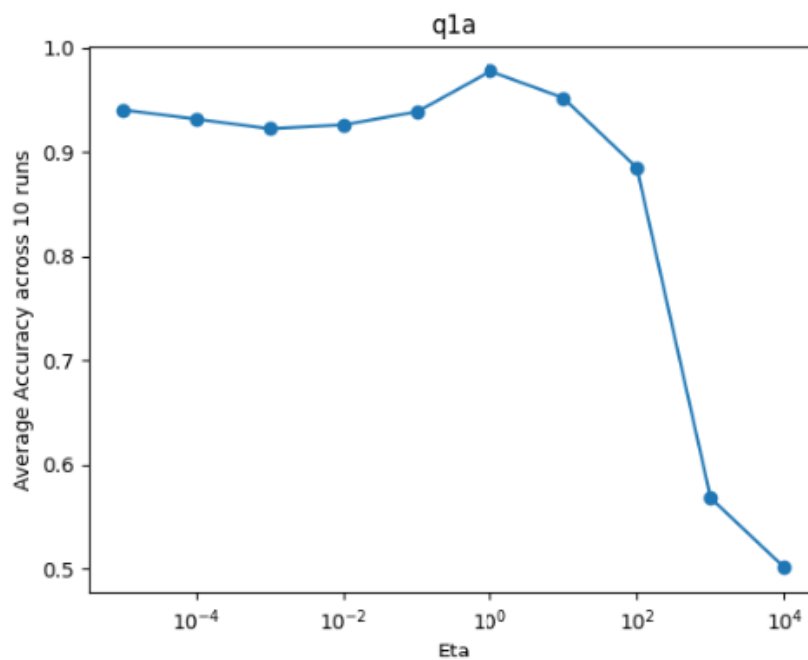


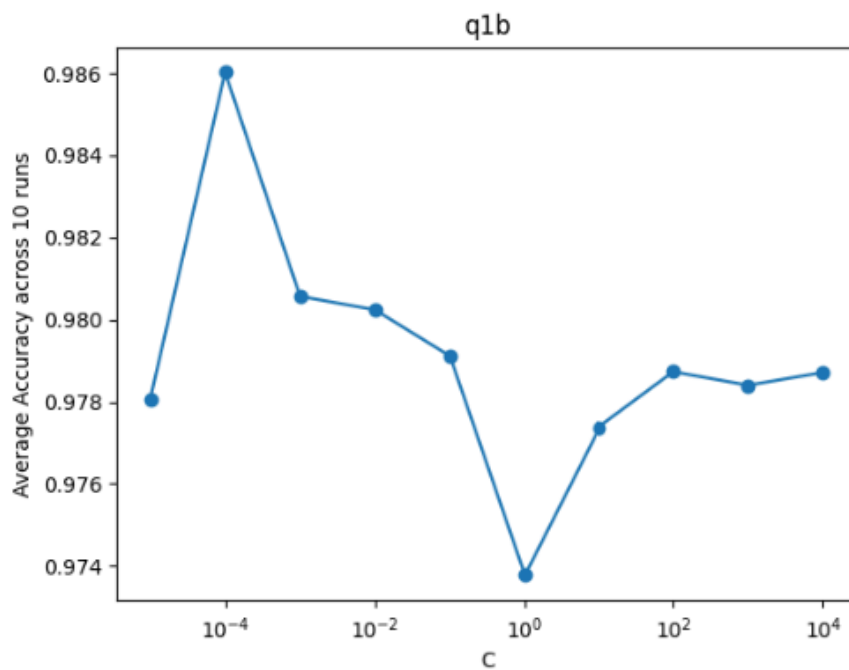
חלק מעשי

שאלה 1

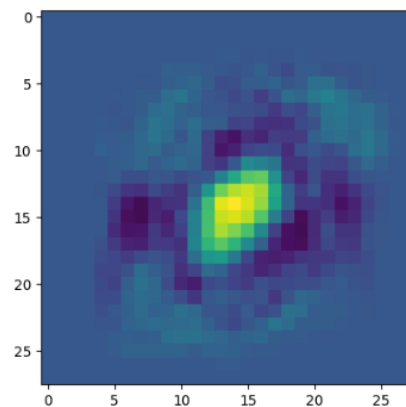
א. מקבלים כי ה η הכי טוב הוא $10^0 = 1$



ב. מקבלים כי ה C הכי טוב הוא $10^{-4} = 0.0001$



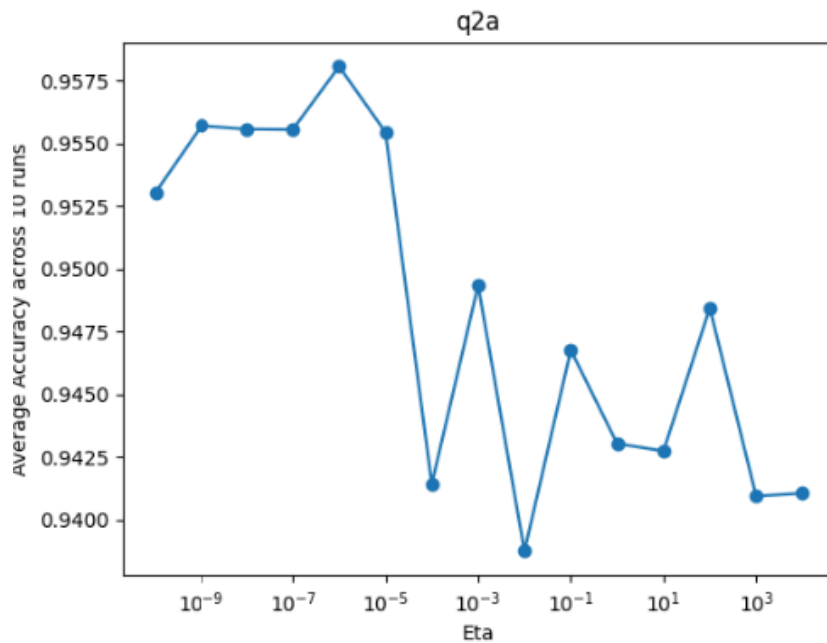
ג. התמונה מנרמלת את וקטור w , האזורים הכהים יותר מציינים סיכוי גבוה יותר לקבל 0 ואילו הבהירים יותר (במרכז) מציינים סיכוי גבוה יותר לקבל 8, ככל שהצבע בהיר המשקל גדול יותר. לכן ההבדל העיקר ביניהם הוא באמצע.



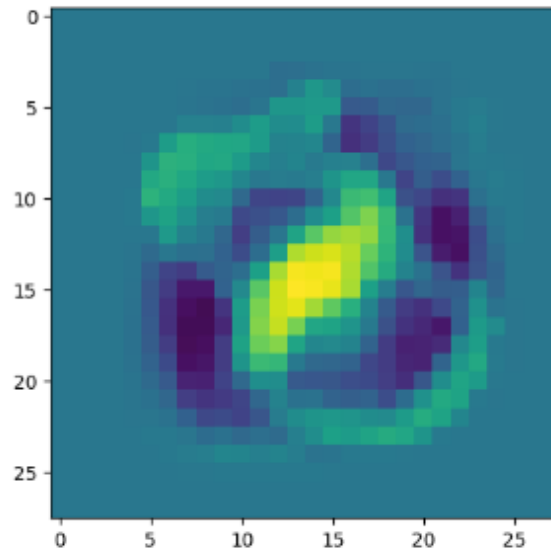
ד. מקבלים כי הדיוק הטוב ביותר הוא 0.9912998976458547.

שאלה 2

א. מקבלים כי ה η הכי טוב הוא 10^{-6}



ב. מקבלים כי הדיוק הטוב ביותר הוא 0.9616171954964176.



ג. מתחילים עם $w=0$ לכן בהתחלה הנורמה היא 0. ה η כיוון שהיא מנורמלת לפי t היא קטנה עם הזמן ולכן קצב הגידול של w קטן עם הזמן. לכן בהתחלה אנחנו רחוקים מה w האופטימלי אבל עם התקדמות באיטרציות של ה SGD ובגלל הרנדומליות נצפה להתקרב למינימום.

