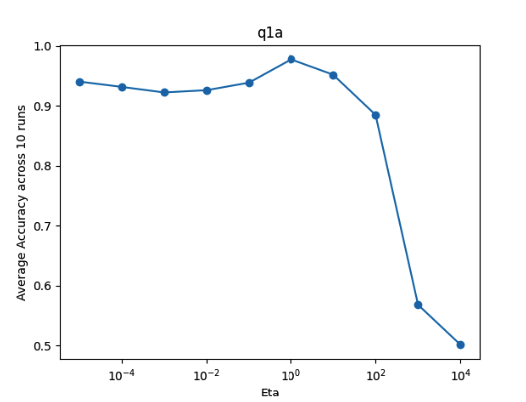
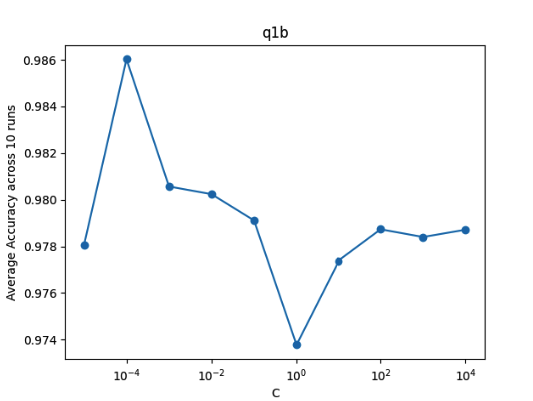
**חלק מעשי**

**שאלה 1**

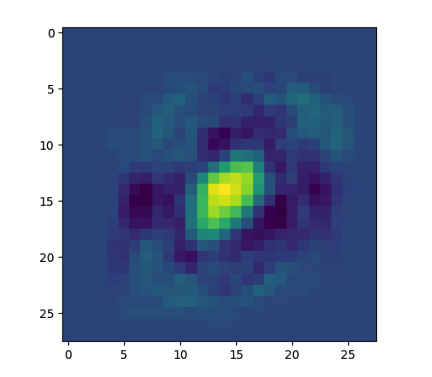
1. מקבלים כי ה eta הכי טוב הוא



1. מקבלים כי ה C הכי טוב הוא



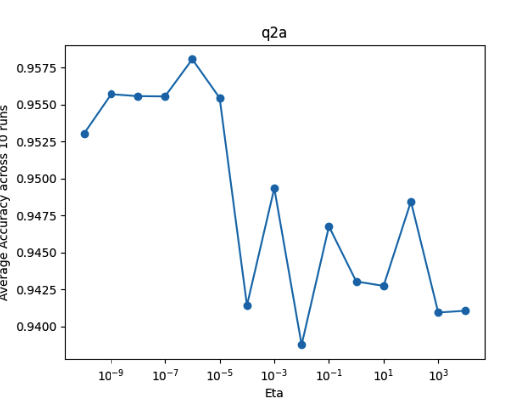
1. התמונה מנרמלת את וקטור w, האזורים הכהים יותר מציינים סיכוי גבוה יותר לקבל 0 ואילו הבהירים יותר (במרכז) מציינים סיכוי גבוה יותר לקבל 8, ככל שהצבע בהיר המשקל גדול יותר. לכן ההבדל העיקר ביניהם הוא באמצע.



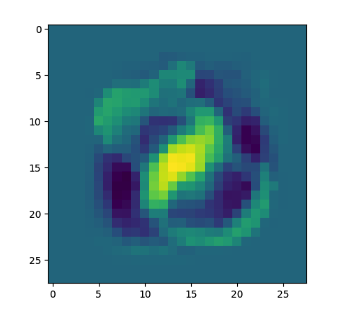
1. מקבלים כי הדיוק הטוב ביותר הוא 0.9912998976458547.

**שאלה 2**

1. מקבלים כי ה eta הכי טוב הוא



1. מקבלים כי הדיוק הטוב ביותר הוא 0.9616171954964176.



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

