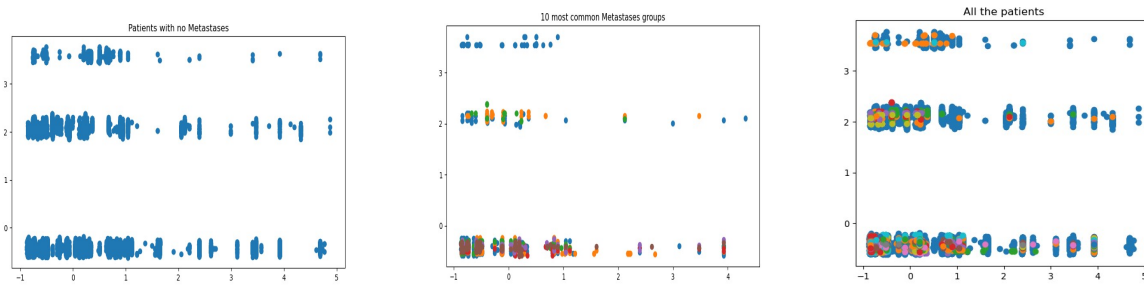


משימה 3

כפי שניתן לראות לפי ה-PCA של הדאטה שלנו לא מצאנו אף אפיון מיוחד או הפרדה בין החולים עם/בלי הגרורות. הרצנו 4 סוגים של PCA. עבור כל ה-train, רק דגימות בלי גרורות, רק דגימות שלהן יש גרורות, ועבור הדגימות השייכות ל-10 הקבוצות הגדולות ביותר של חולים עם גרורות. לא הצלחנו למצוא הפרדה בין קבוצות ולכן לא השתמשנו בזה לטובת הפרדיקציה.

כן ניתן לשים לב שיש קבוצות שמופיעות יותר בערכים מסויימים, לדוגמה אנו רואים בשרטוט המרכזי שכל הדגימות השייכות לקבוצה בחומה מקבלות ערכים שליליים. אם זאת זה לא מספיק בשביל להפריד בין הקבוצות.



בנוסף, ניסינו ללמוד על הדאטה בצורות לא מפותחות נוספות: השתמשנו באלגוריתם Nearest Neighbors ו-KMean Clustering על תתי קבוצות שונות של נתונים בדאטה. הגרפים שנבאו מהמחקר היו ללא הפרדה ברורה בין הקבוצות השונות והיה קשה להניב מהם מידע שימושי.