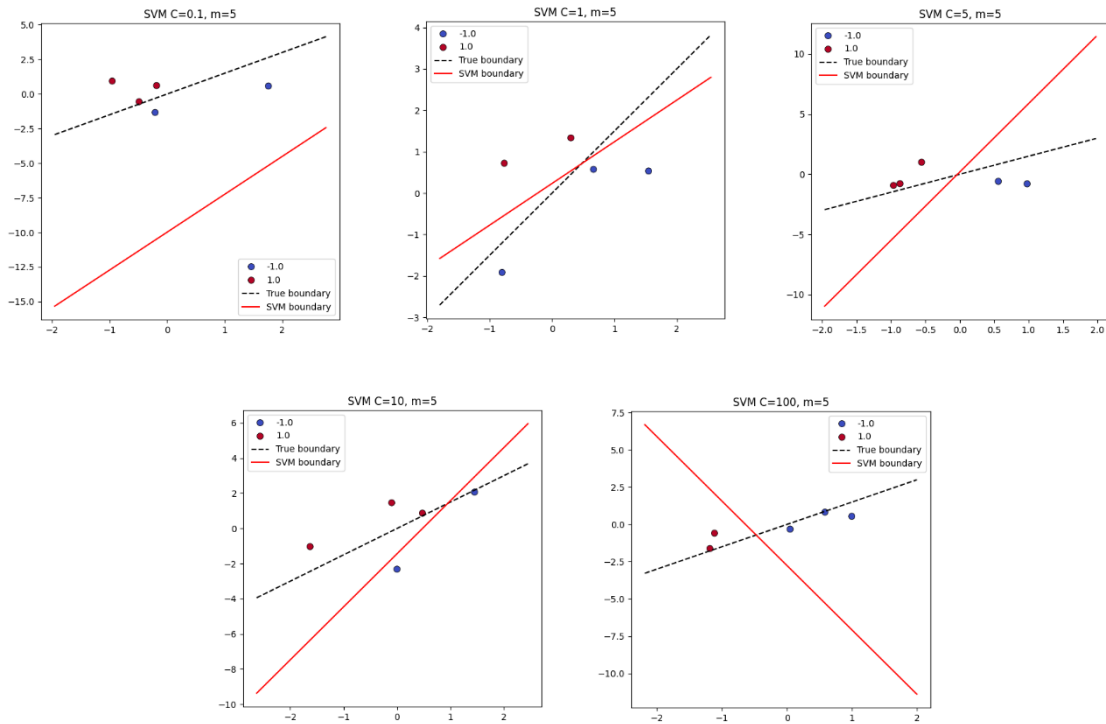


## Practical Part:

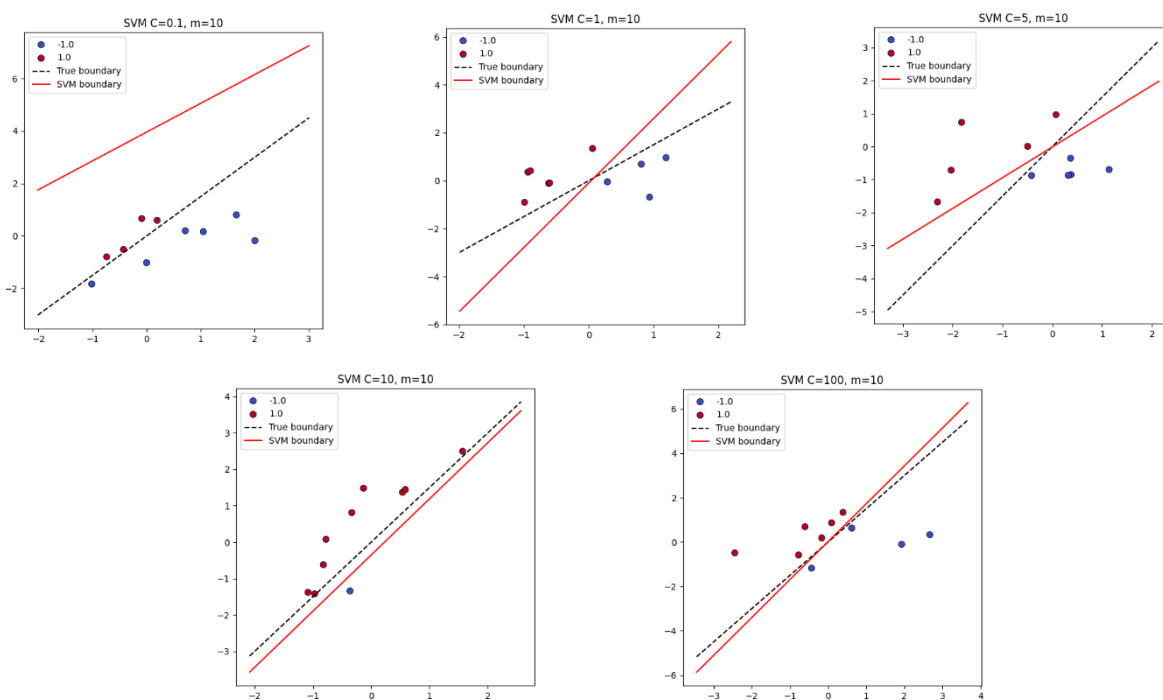
### SVM:

1)

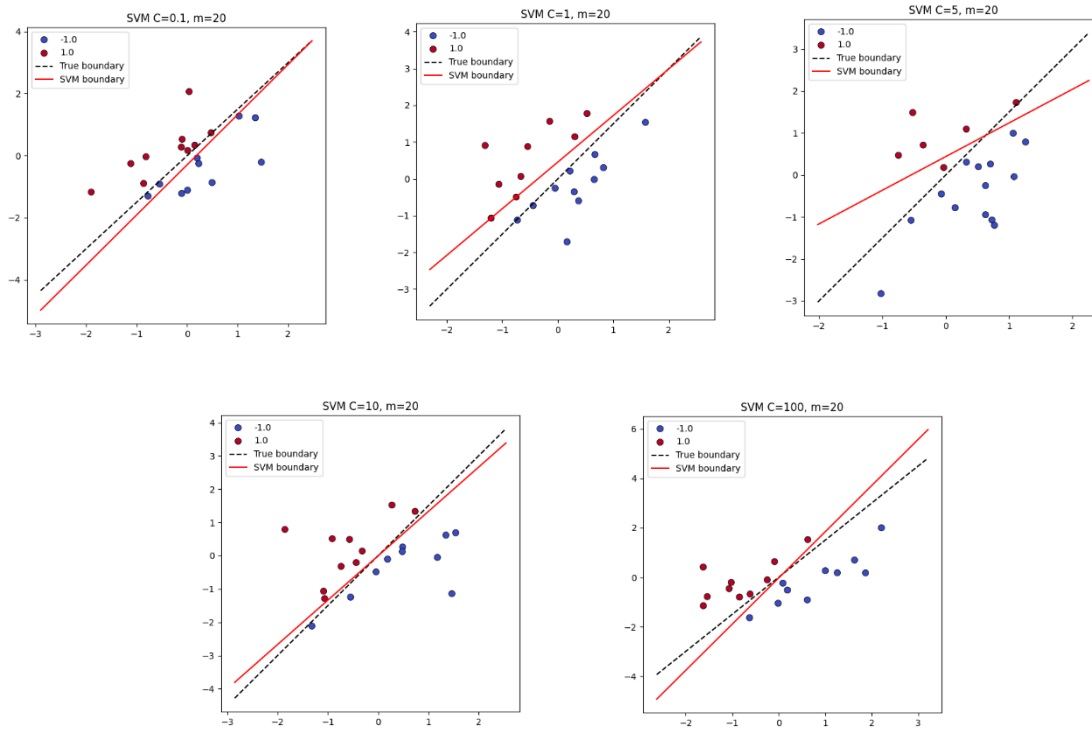
M = 5:



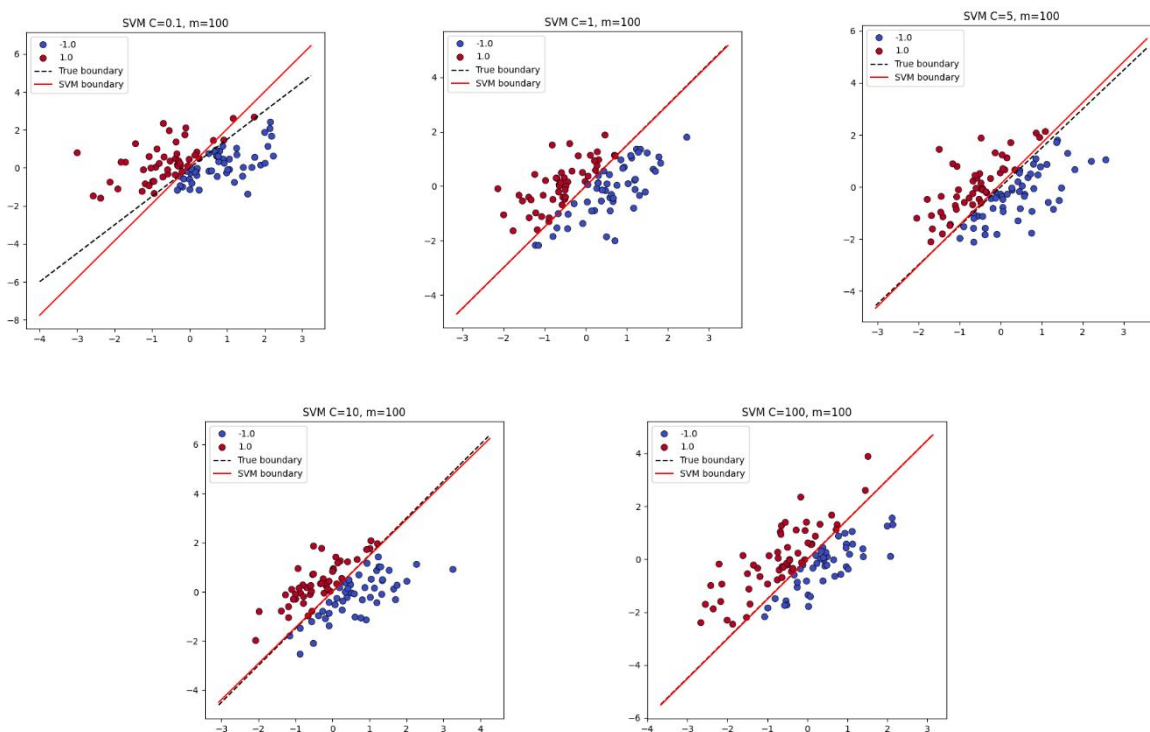
M = 10:



M = 20:



M = 100:

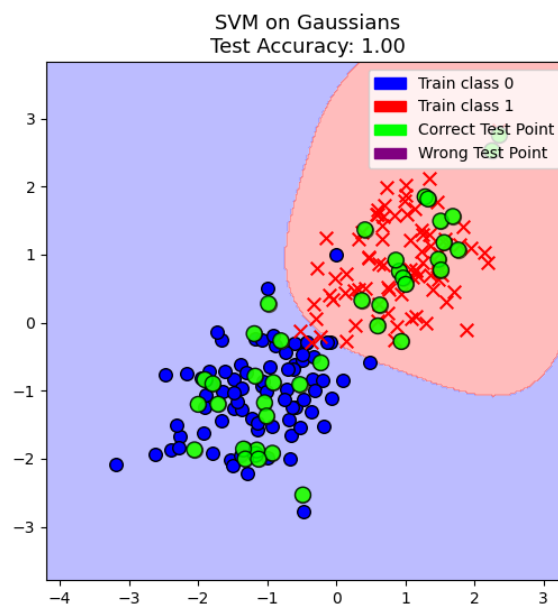
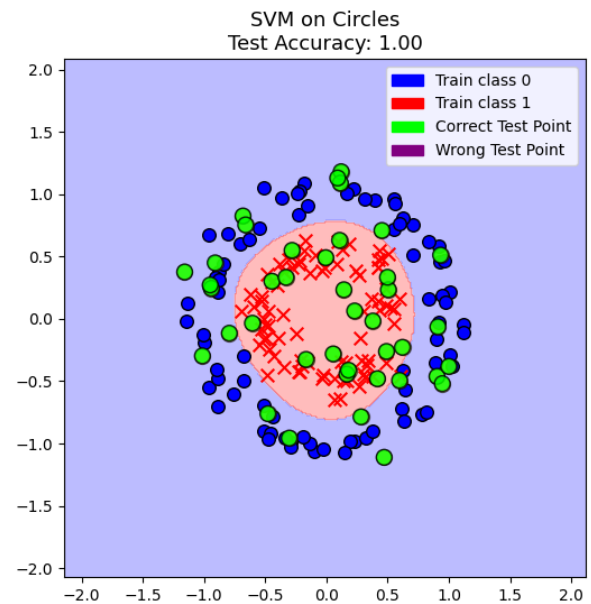
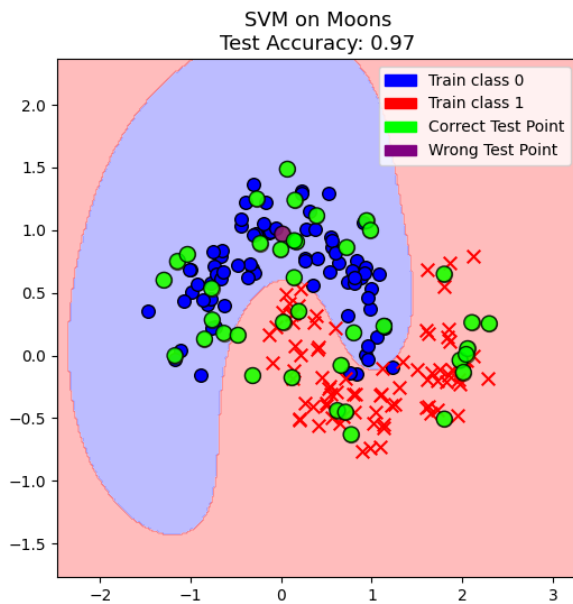


(2) ב-SVM ערך ה- $C$  מייצג כמה "עונש" אנו נותנים על כל טעות של המודל. כלומר ככל שה- $C$  יותר קטן המודל מאפשר יותר טעויות על מנת ליצור את המרווח הגדול ביותר. זה תואם את הגרפים שקיבלנו. ניתן לראות שככל שה- $C$  יותר גדול הקו עושה פחות טעויות בסיווג ומעביר קו עם מרווח קטן יותר מהדגימה הקרובה ביותר אליו. ובהתאמה ככל שה- $C$  קטן יותר הסיווג עושה יותר טעויות אך ממקסם את המרווח.

(3) לפי עקרונות למידת PAC ככל שיש יותר דגימות כלומר  $M$  גדל מוביל לכך שאפסילון (מרווח הטעות) יקטן בהסתברות גבוהה יותר. זה מתבטא באמצעות סיבוכיות הדגימה שאומרת כמה דגימות צריך על מנת להבטיח שגיאה קטנה מאפסילון. ואכן נשים לב מהגרפים שקיבלנו שככל ש- $M$  קטן יותר מודל ה-SVM יוצר סיווג פחות מדויק עם רעש שככל הנראה יחזיר טעויות על העולם האמיתי בהסתברות די גבוהה (ניתן לראות שקו הסיווג שהמודל העביר המסומן באדום רחוק מאוד מקו ההפרדה האמיתי המסומן בשחור). ואכן ככל שהמספר הדגימות עולה כלומר  $M$  גדל הקו שהמודל מעביר מסווג בצורה טובה יותר את הדגימות וככל הנראה יוביל להסתברות הרבה יותר גבוהה למזעור אפסילון (המייצג את טווח הטעות) ניתן לראות זאת גם על ידי הקרבה של קו הסיווג שהמודל העביר לקו המרווח האמיתי.

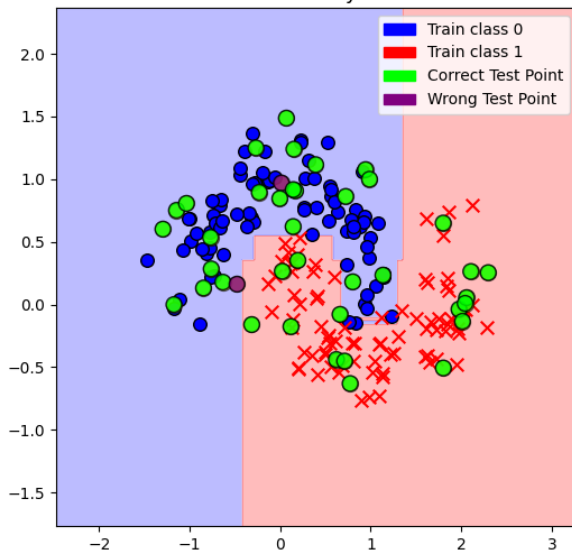
## Classification Boundaries and Model Comparison:

### SVM:

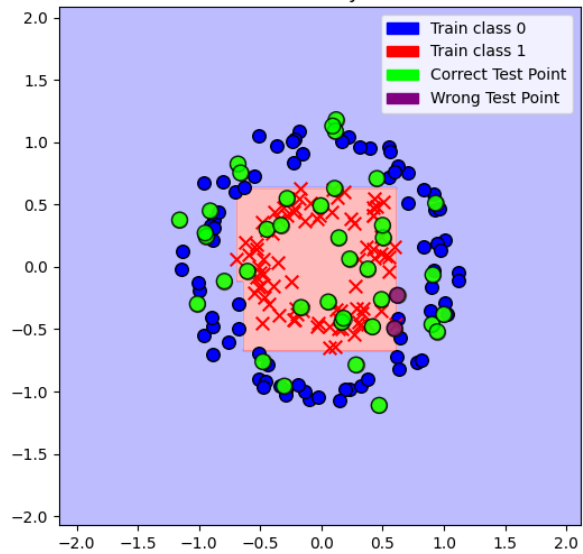


## Decision Tree:

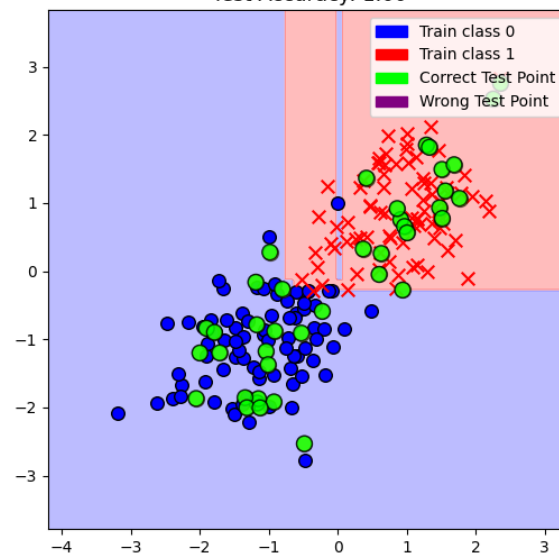
Decision Tree on Moons  
Test Accuracy: 0.95



Decision Tree on Circles  
Test Accuracy: 0.95

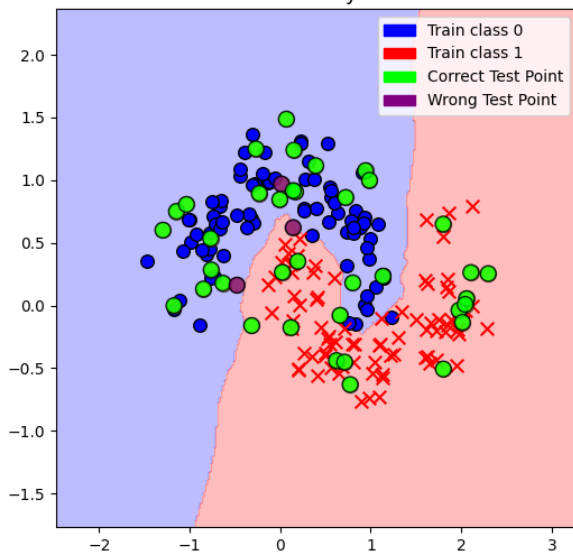


Decision Tree on Gaussians  
Test Accuracy: 1.00

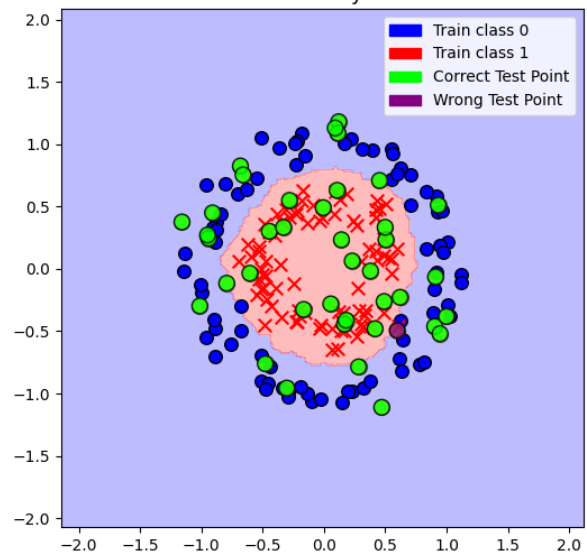


## KNN:

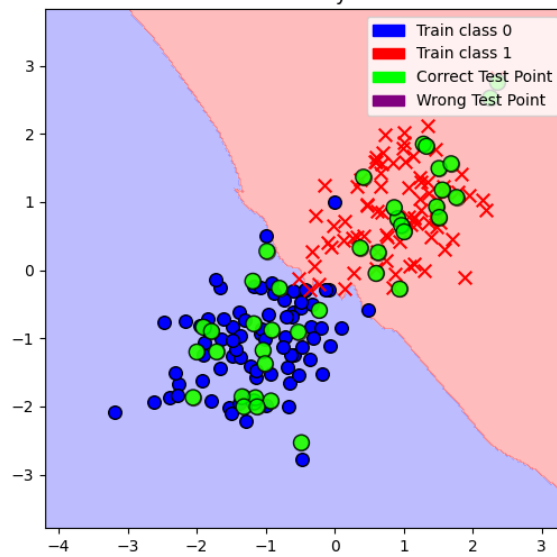
KNN on Moons  
Test Accuracy: 0.93



KNN on Circles  
Test Accuracy: 0.97



KNN on Gaussians  
Test Accuracy: 1.00



2) Moons-דגימות עם הפרדה בצורת ירח- לדאטה סט הזה המודל הטוב ביותר היה SVM אחריו עץ החלטות ולבסוף KNN. תוצאה זאת הגיונית מאחר ולסוג דאטה זה יש הכי הרבה "התערבבות" מרחבית של שני הקלאסים, ולכן הגיוני של KNN יהיה הכי קשה לסווג מאחר ויש הכי הרבה נקודות עם שכנים מהקלאס השני, לגבי עץ החלטות גם הגיוני שהתוצאה תהיה פחות טובה מאחר והוא עובד בשיטת מלבנים והצורה המרחבית של הדאטה אינה מלבנית ולכן הוא יתקשה יותר מ-SVM שיוציא את הסיווג הטוב ביותר מאחר והוא יכול לצייר קלסיפיקציה בצורות דינאמיות שאינן בהכרח ריבועיות.

Circles- דגימות עם הפרדה מעגלית - גם כאן הדאטה סט הטוב ביותר הוא SVM וזה הגיוני שוב בשל היכולת שלו לייצר קלסיפיקציה דינאמית שאינה קווים ישרים כמו עץ החלטות שאכן הוציא את הסיווג הגרוע ביותר. לגבי KNN מדובר בסיווג טוב יחסית אך פחות מ-SVM גם כן בשל היכולת של KNN לייצר צורות דינאמיות יותר.

Gaussians- שלושת סוגי הסיווג קיבלו פה תוצאה אופטימלית וזאת בשל העובדה שהדאטה מסודר בצורה נוחה לשלושת המודלים להוציא תוצאה אופטימלית, יש הפרדה וקרבה מספקת בין הדגימות כך שכל שלושת המודלים מצליחים להוציא סיווג אופטימלי.

3) SVM- האלגוריתם מייצר גבול גמיש וחלק בצורות דינאמיות שאינן בהכרח לינאריות תוך ניסיון למקסם את המרחק של הקו משתי המחלקות (הדגימות הקרובות ביותר מכל מחלקה).

עץ החלטות- האלגוריתם יוצר מלבנים לפי שני הצירים ועובר על הדגימות תוך חלוקה של המרחב לשניים בצורה רקורסיבית וסיווג על ידי רוב במלבן שנוצר עד לעצירה על ידי קביעת פרמטר שמסמן מתי לעצור את החלוקה. אלגוריתם זה מחלק את המרחב ל"קופסאות" בצורת מלבן שלכל אחת סיווג אחר. עובד טוב כשהדאטה "מרובע" באופי.

KNN – האלגוריתם מקבל פרמטר K ומסווג כל נקודה לפי סיווג K השכנים הקרובים אליה ביותר ולכן הוא יוצר גבולות מאוד גמישים ובצורות לא סדירות (לעיתים סוג של מובלעות) בהתאם לדאטה עצמו.