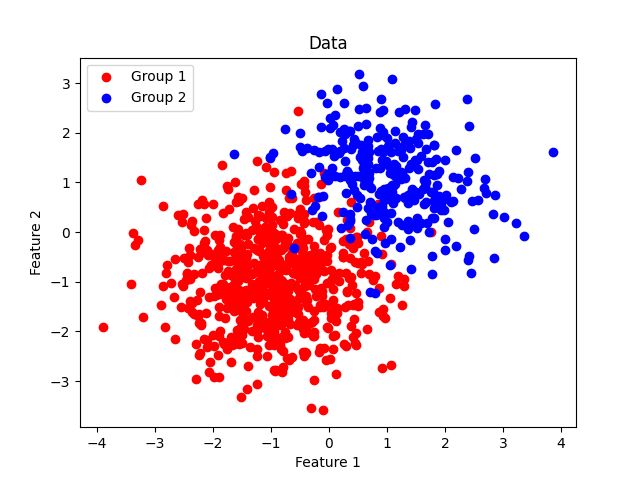
דוח למידה שאינה מוכוונת

1. שאלה 1

* חזינו את הנתונים,ייצגנו אותם על ידי תמונה בגרף. הממוצעים של שתי הקבוצות.על ידי שימוש בפונקציה np.random.multivariate\_normal, ניתן ליצור נקודות נתונים אקראיות בהתבסס על הממוצעים ומטריצות הפצה הנתונות עבור כל קבוצה. זה מאפשר ליצור קבוצת נתונים עם 1000 נקודות נתונים בסך הכול.
* על ידי ציור הקבוצות הנתונים הללו על גרף, נוכל לראות את ההפצה וההפרדה של שתי הקבוצות. כל קבוצה תציג את צורתה והתפשטותה המאפיינות על פי הממוצעים ומטריצות הפצה שסופקו



1. שאלה 2

* אנחנו בחרנו לממש אצ אלגוריתם GMM
* האלגוריתם GMM עובר בשני שלבים: EXPECTATION AND MAXIMIZATON
* בשלב (E-step), האלגוריתם מחשב את התחזיות (responsibilities) עבור כל נקודת נתונים וכל קלסטר. התחזיות מייצגות את הסיכוי של כל נקודת נתונים להתייחס לקלסטר המתאים. התחזיות נחשבות באמצעות הפרמטרים הנוכחיים של הממוצעים, מטריצות הפצה וקואפיציאנטיות, ובעזרת פונקציית הגאוסית המותאמת (gaussian function).לאחר מכן, התחזיות נורמליזן כך שיחד מתייחסות לכל נקודות הנתונים יכולה להיות 1.
* בשלב (M-step), הפרמטרים מתעדכנים בהתאם לתחזיות הנוכחיות. עבור כל קלסטר, הממוצע מתעדכן כממוצע מוקפץ של נקודות הנתונים, מטריצת הפצה מתעדכנת כמטריצת קובריאנס מוקפצת של נקודות הנתונים.
* התהליך נמשך עד שהמספר המרבי של צעדים מקסימיזציה מוגבל על ידי הספק המקסימלי של האיטרציות (max\_iterations).
* לאחר התקבלות על התחזיות המתקרבות למספרים קבועים או לאחר סיום מספר האיטרציות המרבי, **האלגוריתם נחשב להתכנס**. זה אומר שהפרמטרים התקבלו בערך קבוע, ואין שינויים ניכרים ביניהם בין איטרציה לאיטרציה.
* במקרה של קוד ה-GMM שמימשנו, ניתן לעקוב אחרי מספר האיטרציות שבוצעו (iterations) ולבדוק אם הם עברו את המספר המרבי שהוגדר ל-100. אם האיטרציות הגיעו למספר המרבי, ניתן להסיק כי האלגוריתם לא התכנס בהתאם להגדרה, ויתכן שצריך לשנות את הגבול של המספר המרבי או להגדיר תנאי התכניתות נוסף כדי להגיע להתכנסות תוך מסגרת הזמן הנדרשה.
* אחרי כמה בדיקות ראינו שהאלגוריתם לא מתכנס מתחת ל 91 איטרציות רק מ 92 ומעלה לפני המקרה שלנו.

