

תשובות:

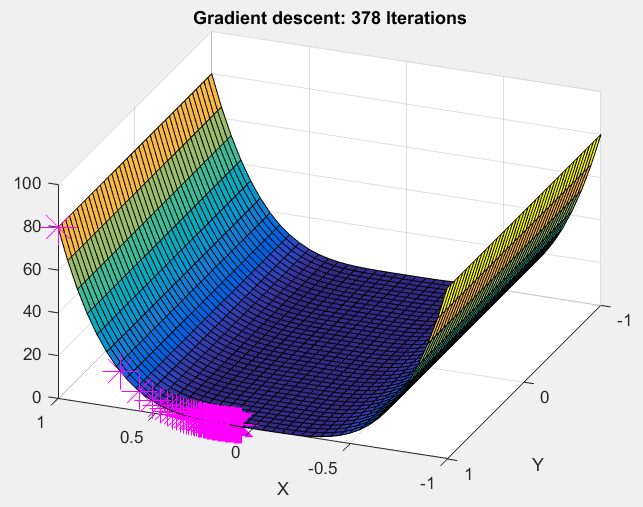
1. לפונקציה נתונה יש נקודת מינימום ב- (0,0) וערך הפונקציה הוא 0.

Figure 1 - Gradient Descent algorithm with regular step function

1. עבור ערך צעד זה האלגוריתם אינו מתכנס.
2. נריץ את האלגוריתם, ונציג את התוצאות:

ניתן לראות כי האלגוריתם מתכנס תוך 378 איטרציות לרמת דיוק 1e-4. נשים לב שההתכנסות בציר X הרבה יותר מהירה, כי ערך הנגזרת בכיוון X גדול ולכן גודל הצעד בכיוון X משמעותית יותר גדול מהצעד בכיוון Y. ערך הנגזרת בכיוון Y מאוד קטן ולכן אנחנו מגיעים לדיוק הרצוי הרבה לפני שמגיעים לנקודת מינימום.

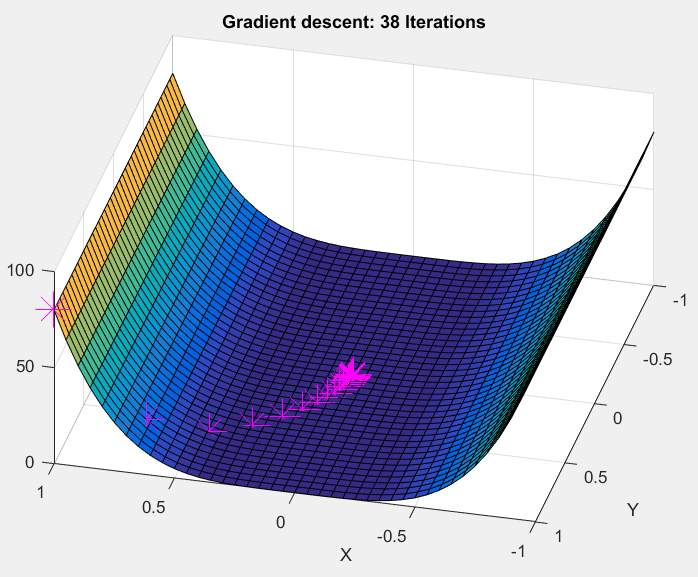
1. נריץ את האלגוריתם תוך שימוש בצעד ניוטון ונציג את התוצאות. נשים לב כי הפעם ההתכנסות הרבה יותר מהירה ותוך 38 איטרציות אנחנו מתכנסים לנקודת מינימום של הפונקציה.

Figure 2 - Gradient Descent algorithm with Newton step function

יתרונות צעד ניוטון:

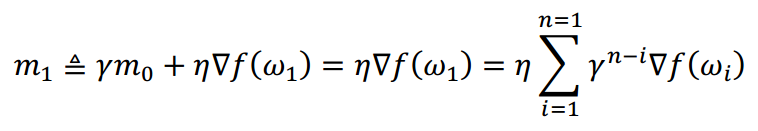
* אין צורך לבחור גודל צעד. הכפל במטריצה הפכית של הסיאן מתאים את גודל הצעד אוטומטית לכל ציר. למשל, בכיוון בו ערך הנגזרת החלקית קטן, גודל הצעד יהיה גדול והפוך.
* ההתכנסות הרבה יותר מהירה. בזכות התקדמות מהירה בכל הצירים, האלגוריתם מתכנס לנקודת מינימום יותר מהר.

חסרונות צעד ניוטון:

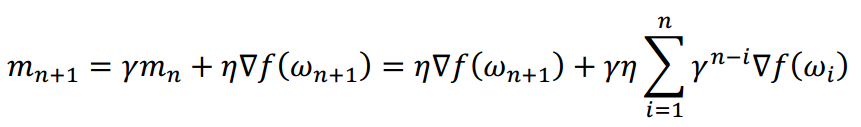
* יש צורך להפוך את מטריצת ההסיאן בכל איטרציה. פעולה זאת כבדה חישובית.

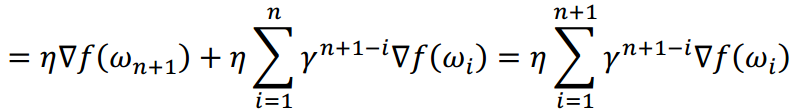
1. נוכיח באינדוקציה:

בסיס: n=1. לכן מתקיים עבור m0=0



הנחה: נניח שמתקיים עבור n כלשהו. נוכיח כי מתקיים עבור n+1:





מ.ש.ל

משמעות הביטוי היא שנשמרת "היסטוריה" של ערכי הגרדיאנטים של נקודות שביקרנו בהם בעבר, כך שהמשקל קטן יותר עבור נקודות ישנות יותר.

1. נריץ את האלגוריתם ונציג את התוצאות: קיבלנו התכנסות לרמת דיוק נדרשת תוך 713 איטרציות.

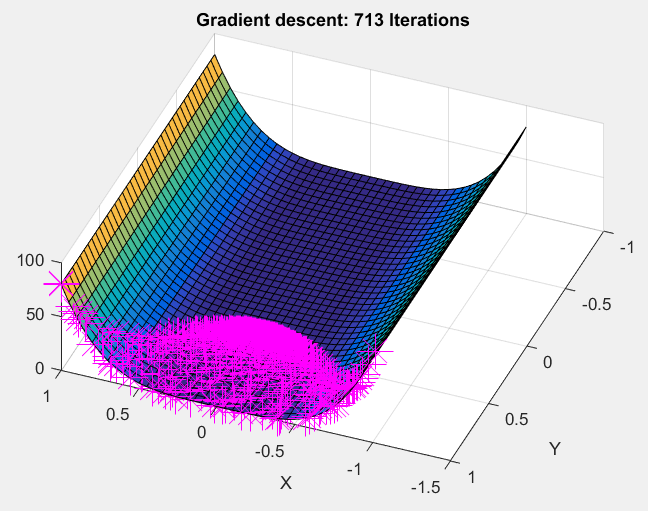


Figure 3 - Gradient Descent algorithm with Momentum step function