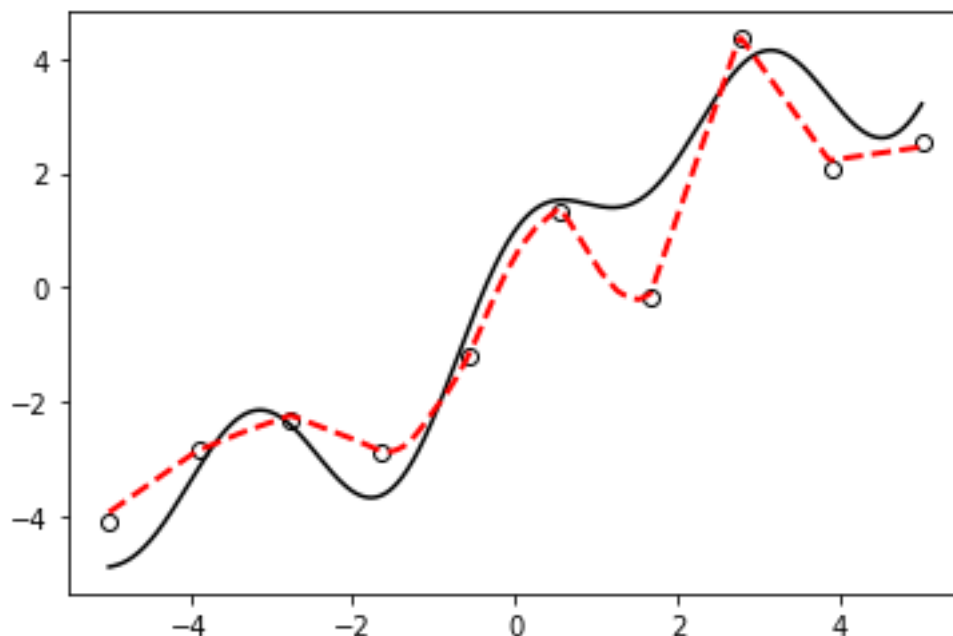


## התאמת יתר

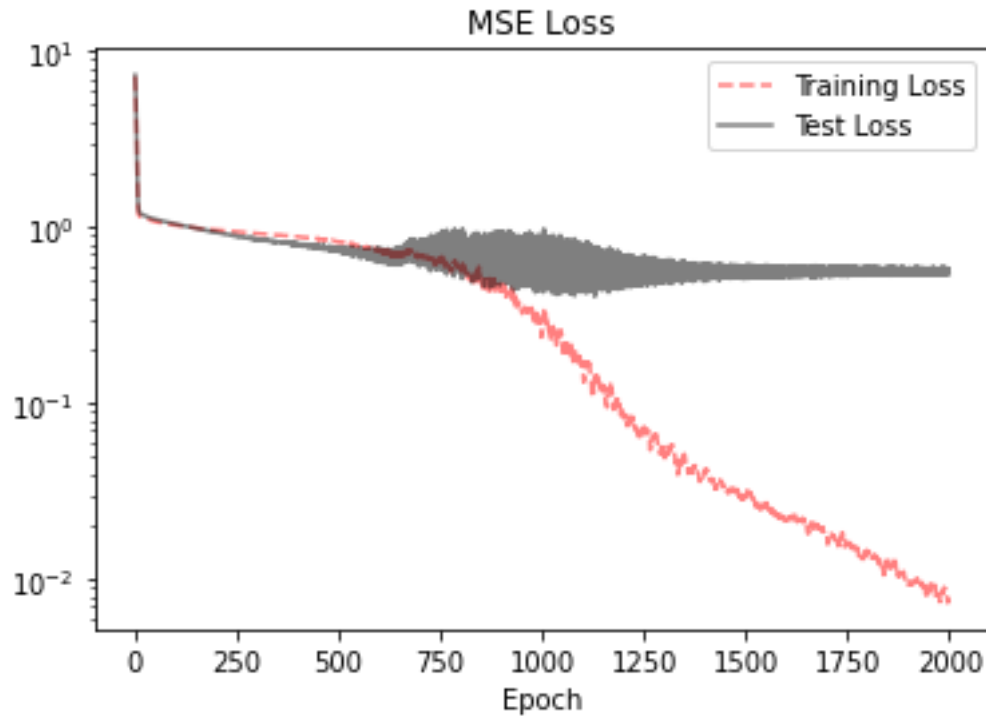
את הפרק הקודם סיימנו באימון רשת נוירונים בסיסית לפתרון בעיית רגרסיה פשוטה וראינו שהתוצאה אשר התקבלה שחזרה בקווים כללים את הכלל האמיתי אשר ממנו נדגם סט האימון, אך בוודאי שלא את כל המורכבות אשר באה לידי ביטוי בו. זהו בדרך כלל אינו המצב כאשר אנו מאמנים רשתות נוירונים מודרניות. בהיותן רשתות עמוקות בעלות מספר עצום של פרמטרים, הן יכולות לבטא מורכבות כמעט בלתי מוגבלת. לצורך הדגמה, נאמן את הרשת המוגדרת בקטע הקוד הבא, רשת מורכבת יותר, אך עדיין קטנה ביחס לרשתות מודרניות, לפתרון אותה הבעיה. כתוצאה נקבל פונקציה המתאימה לסט האימון באופן מושלם, כפי שניתן לראות באיור העוקב.

```
model=nn.Sequential(nn.Linear(1,100),  
                    nn.ReLU(),  
                    nn.Linear(100,100),  
                    nn.ReLU(),  
                    nn.Linear(100,100),  
                    nn.ReLU(),  
                    nn.Linear(100,1))
```



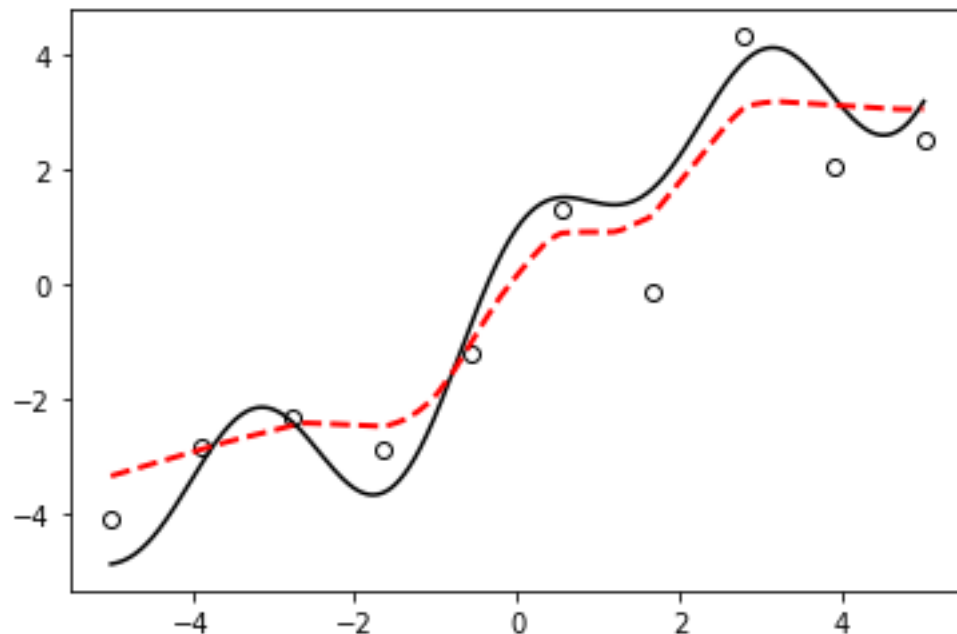
תוצאה זו לא הייתה בעייתית לו סט האימון היה מייצג באופן מושלם את הכלל האמיתי (לו הנקודות היו יושבות בדיוק על הקו השחור), אך זהו אינו המצב, שכן סט האימון נדגם עם רעש. כמו בדוגמה, כך גם במציאות, סט האימון בו אנו משתמשים לאימון כל מודל מכיל מידע רב על הכלל אותו אנו רוצים ללמד את הרשת, כלל הרלוונטי גם לדוגמאות מחוץ לסט האימון. יחד עם זאת, סט האימון מכיל גם מידע לא רלוונטי, הנובע מתהליך איסוף הנתונים הרועש, ומגודלו הסופי של הסט. אלגוריתם למידה מוצלח ילמד את המידע הרלבנטי לכלל הנתונים, ויתעלם מהמידע הספציפי לסט האימון – תהליך הנקרא "הכללה" (Generalization).

מטרתנו בפרק זה היא אם כן, להוסיף לתהליך אימון הרשת כלים אשר יפחיתו את שגיאת ההכללה (Generalization error), אותה נאמדם בעזרת סט הבדיקה של הנתונים – אוסף נתונים נוסף אשר לא השתמשנו בו בעת האימון. רשת מוצלחת אשר ביצעה הכללה ולא "שיננה" את נתוני האימון תפגין ביצועים טובים גם על סט נתונים זה. המצב ההפוך, בו הרשת משננת את סט האימון בא לידי ביטוי בכך שבמהלך תהליך האימון פונקציית המחר של סט האימון ממשיכה לרדת, אך כאשר אנו מחשבים את המחר של סט הבדיקה – השגיאה אינה משתנה, ואף עולה לאורך דורות האימון. מצב זה, הנקרא התאמת יתר (Overfitting), ניכר בבירור בדוגמה הנ"ל: ראו את גרף שגיאת האימון ושגיאת הבדיקה המאזין להלן.



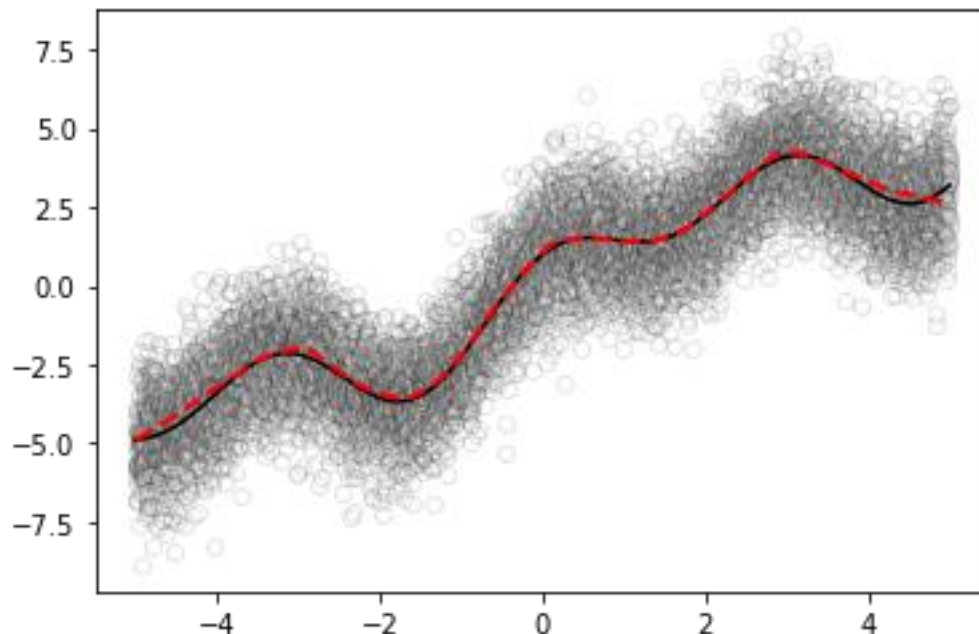
מאיר זה ניכר שהחל מה-Epoch-500 לערך, הרשת התחילה לשנן את סט הנתונים – היא המשיכה להתקרב אל הנקודות הלבנות, שכן שגיאת האימון המשיכה לרדת, בעוד שהשגיאה על סט הבדיקה לא ירדה. הקרבה אל הנקודות הלבנות באה על חשבון הקרבה למודל התיאורטי (הקו השחור) ומכך אנו רוצים להימנע.

שינוי תהליך אימון הרשת בכדי להקטין את שגיאת ההכללה נקרא רגולריזציה (Regularization), ובהמשך היחידה נלמד מספר שיטות שונות לבצע זאת. הראשונה שניתן להפעיל היא הבסיסית ביותר: אם החל מ-Epoch מסויים שגיאת האימון מתחילה לעלות, אז נאמן את הרשת עד ל-Epoch זה בלבד. באופן לא מפתיע שם השיטה הוא עצירה מוקדמת (Early Stopping). תוצאת אימון הרשת הנ"ל לאורך 500 דורות מאויירת להלן, וממנה ניכר בבירור שהרשת עוד לא הספיקה לשנן את הנתונים.



כאשר אנו משתמשים בשיטה זו, יש לתת את הדעת לכך שסט הבדיקה שימש לבחירת הדור האידיאלי לעצירת תהליך האימון. על כן, שגיאת הבדיקה הנמדדת לפי סט זה אינה בהכרח תייצג נאמנה את שגיאת ההכללה – הרי דור העצירה נבחר בדיוק כדי למזער שגיאה זו. בהתאם, אם ברצוננו לאמוד את שגיאת ההכללה של הרשת יש להשתמש בסט נתונים נוסף, שלא השתמשנו בו כלל. באופן פורמלי, השם "סט הבדיקה" שמור לאוסף נתונים נוסף זה, ואוסף הנתונים בו השתמשנו עד עתה לבדיקת ביצועי המודל הוא סט הולידציה (Validation Set). עם זאת, בהמשך, כאשר נשתמש רק בסט אימון וסט נתונים נוסף לבדיקה, נמשיך לקרוא לו סט הבדיקה כמקובל בתחום הלמידה העמוקה, בידענו שאין זה מדויק.

השיטה הבסיסית השניה לרגולריזציה היא הטבעית ביותר – לאסוף יותר נתונים תחת הציפייה שהרעש הבא לידי ביטוי בכל דגימה יהיה בעל משמעות פחותה כאשר סט האימון גדול. ראו לדוגמה, אם נאמן את רשת הרגרסיה הנ"ל על סט אימון גדול הרבה יותר, אשר נדגם באותו תהליך בדיוק כמו סט האימון הקודם, תוצאת הרשת יושבת בדיוק על הקו השחור, כפי שניכר באיור הבא.



ביצועיהן העדיפים של רשתות נוירונים על פני אלגוריתמי למידה אחרים נובעים לא במעט מרעיון פשוט זה – בהיותן כה גמישות, דרוש סט אימון גדול מאוד בכדי לאמנן.

## שאלות לתרגול

- בשאלה זו תיווכחו בקיום תופעת התאמת היתר כפי שהיא באה לידי ביטוי ברשת לסיווג תמונות פרטי הלבוש בסט הנתונים Fashion-MNIST.
  - חלקו את סט נתוני הבדיקה לשני חלקים : ולידציה (80% מסט הבדיקה המקורי) ובדיקה (20% הנותרים).
  - אמנו רשת עמוקה לפתרון בעיית הסיווג עד אשר מתקבלת התאמת יתר לסט הולידציה החדש. זהו זאת לפי גרף שגיאת האימון ושגיאת הולידציה.
  - מצאו את ה-Epoch האידיאלי לעצירת תהליך האימון, ואמנו את הרשת עד Epoch זה.
  - מדדו את ביצועי הרשת לאורך הדורות על סט הבדיקה.
  - הסבירו את הפער בין שגיאת הולידציה לבין שגיאת הבדיקה.
- חזרו על שאלה 1 כאשר סט האימון יהיה 10% בגודלו מסט האימון המקורי. מה תוכלו להגיד על ה-Epoch האידיאלי לעצירה מוקדמת במקרה זה, ביחס לתוצאה הקודמת?
- בשאלה זו נבדוק את ההשפעה של אימון הרשת על סט נתונים אחד, והפעלתו על סט נתונים השונה במהותו.
  - חזרו על שאלה 1 כאשר סט האימון מכיל רק חצי מהמחלקות (מחלקות 0-4), סט הולידציה מכיל רק את המחלקות 0-6 וסט הבדיקה מכיל את המחלקות 5-9. **הערה:** את הדגימות הרלוונטיות תוכלו למצוא בעזרת הפונקציה `torch.where`.
  - האם איסוף נתונים נוספים לסט האימון יעזור להורדת שגיאת ההכללה? הסבירו תשובתכם.