רשתות נוירונים לרגרסיה

בכל הדוגמאות לרשתות נוירונים בהן עסקנו ביחידות הקודמות, פלט הרשת היה הסתברויות חזויות ומטרתנו הייתה לפתור בעיית סיווג למספר מחלקות אך אין להסיק מכך שרשתות נוירונים שימושיות למשימות אלו בלבד. למעשה, ניתן להשתמש בהן לפתרון מספר רב של משימות כגון זיהוי אובייקטים בתמונה, דחיסת מידע, תרגום אוטומטי של קטעי טקסט ועוד רבות אחרות. בעוד שלמטרות שונות קיימות ארכיטקטורות רשת אפקטיביות במיוחד, ועל חלקן נלמד בהמשך הקורס, כל שנדרש באופן בסיסי בכדי לאמן רשת לביצוע משימה שאינה סיווג למחלקות הוא להחליף את "ראש" הרשת, השכבה האחרונה, ואת פונקציית המחיר. בפרק זה נראה כדוגמה כיצד בעזרת שינויים קלים אלו ביכולתנו לפתור בעיית רגרסיה.

נניח כי ברשותנו סט אימון של זוגות מספרים, מהצורה $(x_0, y_0),...,(x_N, y_N)$ ומטרתנו היא לאתר את הקשר הפונקציונלי בין x לבין y, ה"כלל" המחבר בין שני מספרים אלו, כפי שהוא בא לידי ביטוי בנקודות. לצורך המחשה, נניח בהמשך הדוגמה שהכלל המקשר בין שני המשתנים הוא

$$y(x) = x + \sin(x) + \cos(2x),$$

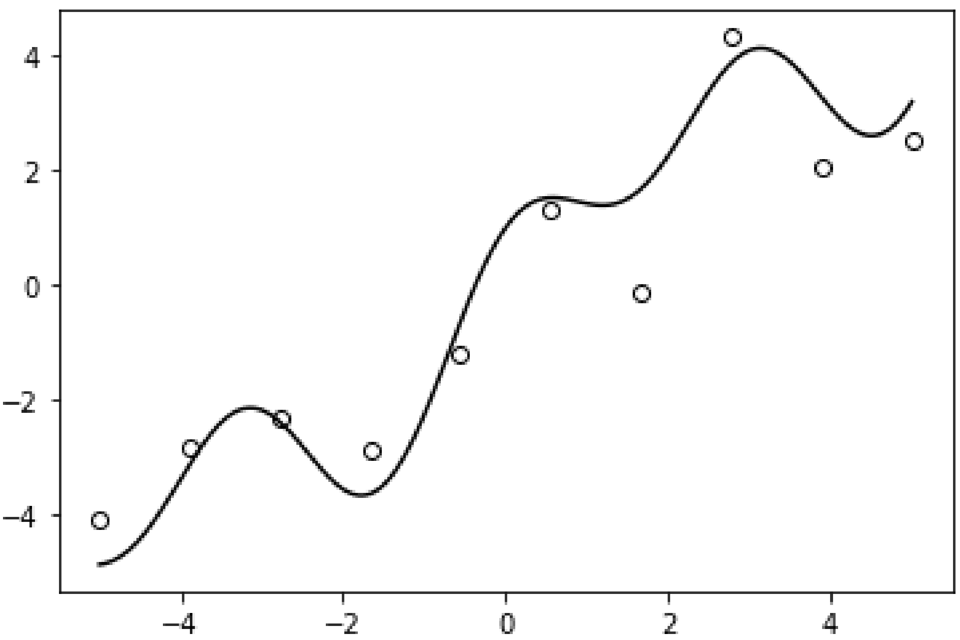
אך איסוף נתוני האימון התבצע בתהליך "רועש", ולכן הנקודות שנדגמו אינן מקיימות כלל זה בדיוק, אלא רק בקירוב. ראו יישום רעיון זה בקטע הקוד הבא.

```
fx= lambda x:torch.sin(x)+torch.cos(2*x) + x
X = torch.linspace(-5,5,1000).resize(1000,1)
Y = fx(X)

Xtrain=torch.linspace(-5,5,10).reshape(10, 1)
torch.manual_seed(1)
Ytrain=fx(Xtrain)+1.2*torch.randn(size=Xtrain.size())

plt.plot(X,Y,color='black');
plt.scatter(Xtrain,Ytrain,color="white",edgecolor="black");
```

פלט:



הקו השחור מייצג את הקשר המתמטי המדויק, בעוד שהנקודות הלבנות את אוסף הנתונים אשר נדגם, בתוספת הרעש. זכרו שבשימושים מציאותיים לא נוכל לצייר את הקו השחור – הקשר בין המשתנים לא יהיה ידוע לנו מראש.

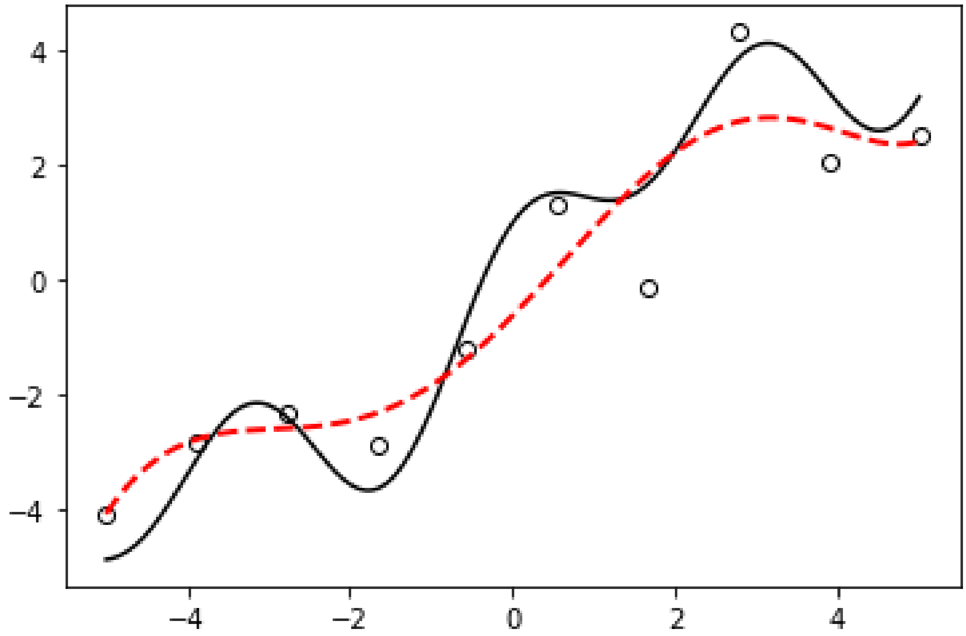
בעיה זו ניתן לפתור, למשל, בעזרת רגרסיה פולינומיאלית: נחפש פונקציה מהצורה

$$y(x)=a_0+a_1x+\ldots+a_nx^n$$

אשר עוברת "הכי קרוב" אל הנקודות הלבנות. לא נוכיח זאת כאן, אך תחת הנחות קלות קיימת בחירה יחידה של פרמטרים $a_0,a_1,\ldots,a_n$ אשר ממזערת את ממוצע ריבועי המרחקים בין התחזית $\left(x_k,y\left(x_k\right)\right)$ לבין נקודת הדגימה מסט האימון $\left(x_k,y_k\right)$. בקצרה, הוקטור

$$\left(a_0',\ldots,a_n'\right)=\underset{\left(a_0,\ldots,a_n\right)\in\mathbb{R}^n}{\arg\min}\frac{1}{N}\sum_{k=0}^{N}\left(y\left(x_k\right)-y_k\right)^2$$

קיים ויחיד. זאת ועוד, ניתן למצוא אותו באופן אנליטי על ידי פתרון מערכת משוואות ליניארית. הפולינום ממעלה 6 המתקבל מהפעלת תהליך זה על אוסף הנקודות שדגמנו לעיל מאוייר להלן.



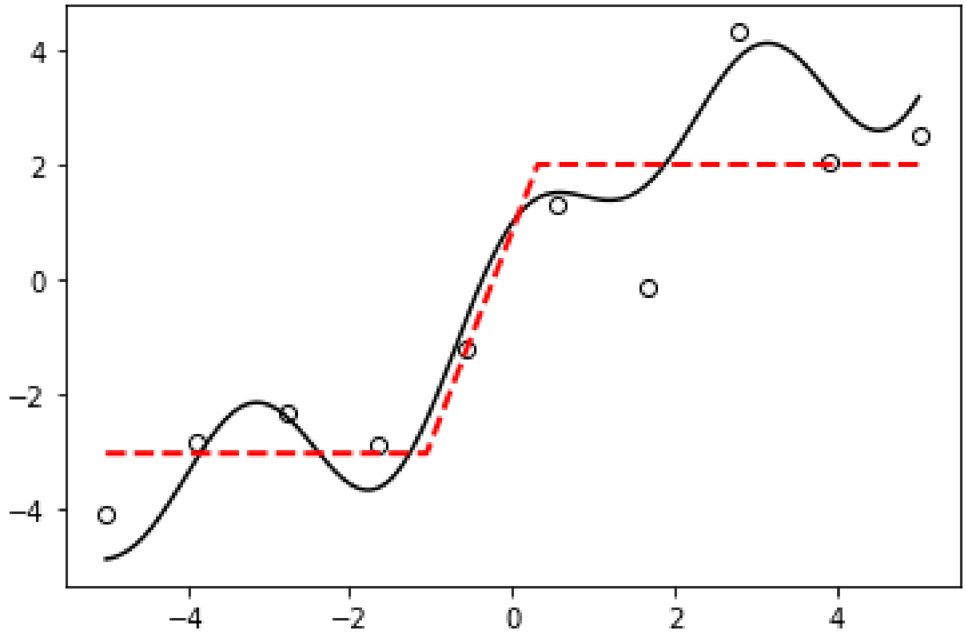
כעת, ברצוננו לשאוב השראה מאלגוריתם זה ולאמן רשת נוירונים לפתור את אותה הבעיה. אם כן, פלט הרשת צריך להיות כפלט הפונקציה $y(x)$, הערך החזוי עבור x נתון. כמו כן יש צורך בפונקציית מחיר חדשה, אך זו למעשה כבר כתובה לעיל – זהו ממוצע ריבועי השגיאות (Mean Squared Error),

$$MSE(\text{Model})=\frac{1}{\#Data}\sum_{\left(x_k,y_k\right)\in Data}\left(\text{Model}\left(x_k\right)-y_k\right)^2.$$

נשים לב שביטוי זה גזיר לפי פלט הרשת, $\text{Model}\left(x_k\right)$, ולכן ניתן להפעיל את האלגוריתם SGD למציאת פרמטרי הרשת אשר ממזערים מחיר זה. נגדיר אם כן רשת אשר השכבה האחרונה שלה היא נוירון ליניארי יחיד, ונתאים את הרשת לנתונים בקטע הקוד הבא.

```
model=nn.Sequential(nn.Linear(1,2),nn.ReLU(),
                    nn.Linear(2,2),nn.ReLU(),
                    nn.Linear(2,1))
optimizer = torch.optim.SGD(model.parameters(), lr=0.1)
loss = nn.MSELoss()
```

כל שנשאר כעת הוא לאמן את הרשת בלולאת אימון סטנדרטית. התוצאה מאויירת להלן.



ניכר שאפשר לשפר את התוצאה, ואכן נעשה זאת בהמשך יחידת הלימוד הנוכחית.

שאלות לתרגול

.1

א. ספרו את מספר הפרמטרים ברשת שהוגדרה לעיל.

ב. נסו להסביר מדוע למרות מספר הפרמטרים הגדול יותר, ניכר שהרשת חוזה פונקציה פשוטה יותר מאשר רגרסיה פולינומיאלית של פולינום ממעלה 6.

2. לעתים בעיית הרגרסיה מנוסחת באופן שקול עם פונקציית המחיר **סכום** הריבועים:

$$S(\text{Model}) = \sum_{(x_k, y_k) \in Data} \left(\text{Model}(x_k) - y_k\right)^2$$

במקום **ממוצע** הריבועים:

$$MSE(\text{Model}) = \frac{1}{\# Data} \sum_{(x_k, y_k) \in Data} \left(\text{Model}(x_k) - y_k\right)^2 .$$

בהנחה שברשותנו הרבה נקודות דגימה, הסבירו את הבעייתיות של שימוש בסכום הריבועים לאימון הרשת. **רמז:** חשבו את הנגזרות של שתי פונקציות המחיר ביחס לפלט המודל.

3. הניחו את הקשר המתמטי הבא בין שלושה משתנים:

$$z = x + y + xy + \sin\left(\frac{x^2}{y}\right) .$$

דגמו סט אימון "רועש" ובצעו רגרסיה בעזרת רשת נוירונים למשתנה z.