

מבוא לבינה מלאכותית – פתרון תרגיל 1:

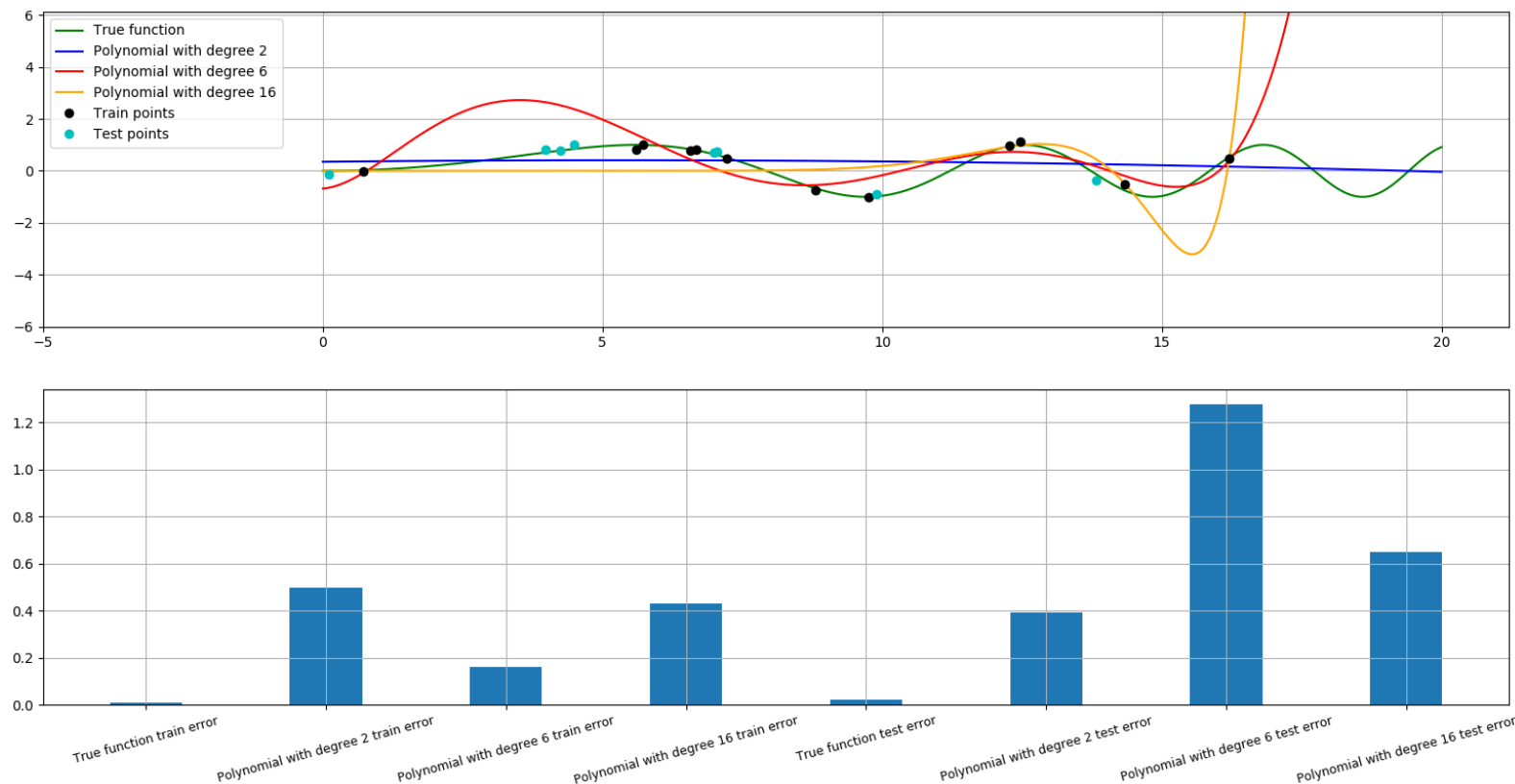
1. נסמן: $a = (a_0, \dots, a_k)$ בתור וקטור המקדמים של הפולינום, $\{(x_i, y_i)\}_{i=1}^n$ התצפיות.
א. ללא רגולריזציה, רוצים למזער את הפונקציה:

$$\text{loss}(a) = \frac{1}{2} \sum_{i=1}^n (y_i - (a_0 + a_1 x_i^1 + \dots + a_k x_i^k))^2$$

ניתן לבדוק ולראות (אם לא ראיתם, אז [קישור](#)) שהמקדמים שממזערים את הפונקציה הזאת הם:

$$a = (X^T X)^{-1} X^T y$$

כאשר y וקטור הערכים y_i , X מטריצת ונדרמונדה (עוד [קישור](#) להסבר).
התוצאה:



במקרה של דרגת פולינום 6 ניתן לראות יפה שיש overfitting – הפולינום מתאים לא רע ל-training, אבל מכיוון שהתאמנו אותו טוב (מדי) ל-training, נקודות חדשות (מה-test) נמצאות רחוק ממה שהערכנו אותו.

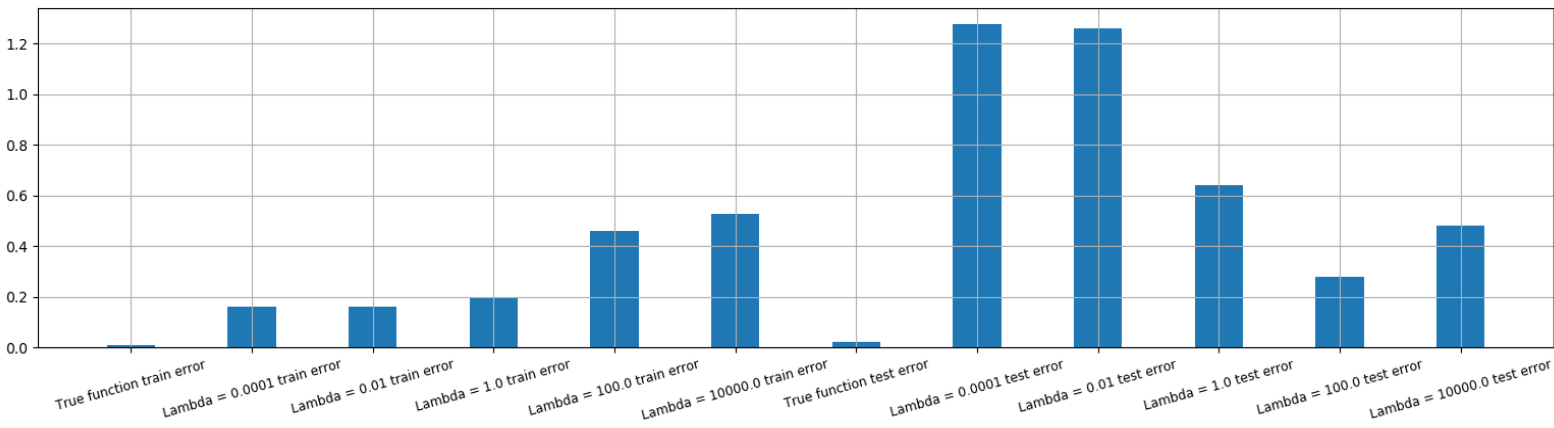
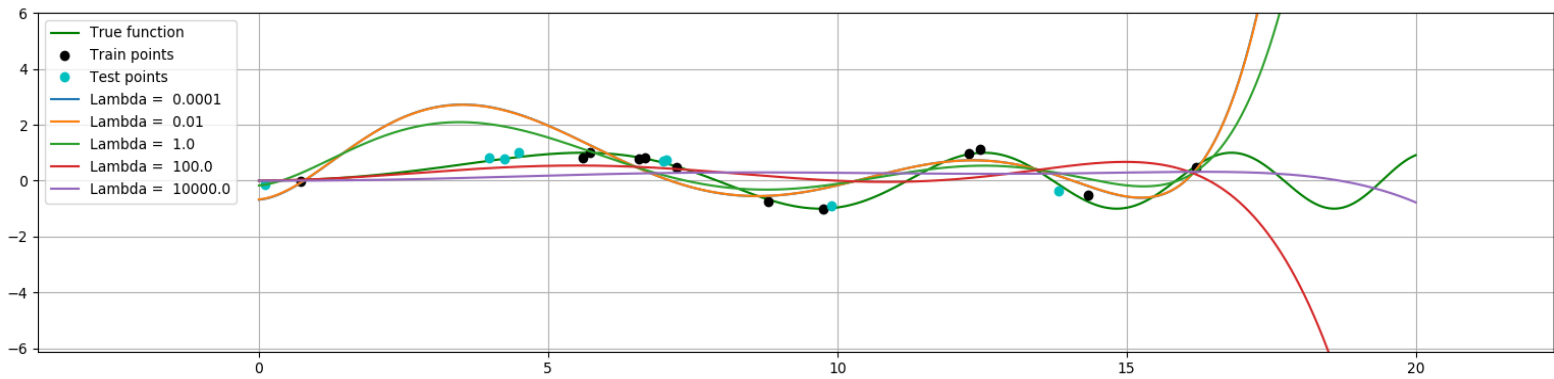
ב. הפעם יש רגולריזציה של ridge, רוצים למזער את הפונקציה:

$$\text{loss}(a) = \frac{1}{2} \sum_{i=1}^n (y_i - (a_0 + a_1 x_i^1 + \dots + a_k x_i^k))^2 + \frac{1}{2} \lambda a^T a$$

ניתן להוכיח שהפעם הנוסחה ל- a נתונה על ידי:

$$a = (X^T X + \lambda I)^{-1} X^T y$$

התוצאות לרגרסיה עבור ערכי λ שונים:



רואים כאן משהו די טיפוסי ל- loss כתלות ברגולריזציה: עבור ערכי רגולריזציה נמוכים, אנחנו כמו בבעיה ללא הרגולריזציה ושוב ב- overfit . עם רגולריזציה מתאימה אפשר ללמוד הרבה יותר טוב, אבל כאשר ניקח את הרגולריזציה להיות גדולה מדי נקבל underfitting , כלומר מצב שבו האילוף שלנו על הפרמטרים מחמיר מדי ולא מאפשר לנו ללמוד.

2. כמו שאמרנו בתרגול, כדי לקבוע שמימד VC של קבוצת פונקציות $\mathcal{H}$ הוא d , צריך למצוא קבוצת נקודות בגודל d ש- $\mathcal{H}$ מנפצת ולהוכיח ש- $\mathcal{H}$ לא מנפצת אף קבוצה בגודל $d + 1$.

א. נראה שהמימד הוא 3:

ניקח למשל את הקבוצה $C = \{1, 2\}$, אז כל הפונקציות שיכולות להיות מ- C לקבוצת התיוגים $\{0, 1\}$ (אלה 4 פונקציות) נמצאות ב- $\mathcal{H}_C$: אפשר לקחת למשל אינטרוולים $(0, 3), (0.5, 1.5), (0, 2), (1, 3)$ ואלה 4 פונקציות שמתאימות ל-4 קומבינציות התיוגים של C .

כעת, תהי $C = \{c_1, c_2, c_3\}$. נניח בה"כ $c_1 \leq c_2 \leq c_3$. אזי, אי אפשר למצוא פונקציה ב- $\mathcal{H}_C$ כזאת ש- $h(c_1) = h(c_3) = 1$ ואילו $h(c_2) = 0$, לכן $\mathcal{H}_C$ לא מנפצת אף קבוצה עם 3 איברים.

ב. לכל קבוצה C מתקיים כי $|\mathcal{H}_C| \leq |\mathcal{H}|$. לכן, אם $|\mathcal{H}| < 2^{|C|}$, אז $\mathcal{H}$ לא תנפץ את C . מכאן ש: $VCdim(\mathcal{H}) \leq \log_2 |\mathcal{H}|$.

ג. נבחר לשם נוחות $C = \{(0, 1), (-0.5, 0), (0.5, 0)\}$. הראו בעצמכם שכל אחת מבין 8 הפונקציות מתקבלת.

מצד שני, לכל C בגודל 4, אי אפשר לנפץ מצבים כאלה (ואחד מהם תמיד מתקבל לכל רביעיית נקודות שנבחר במישור):

