**מבוא לבינה מלאכותית – פתרון תרגיל 1:**

1. נסמן: $a=(a\_{0},…, a\_{k})$ בתור וקטור המקדמים של הפולינום, $\left\{\left(x\_{i}, y\_{i}\right)\right\}\_{i=1}^{n}$ התצפיות.
2. ללא רגולריזציה, רוצים למזער את הפונקציה:

 $loss\left(a\right)=\frac{1}{2}\sum\_{i=1}^{n}(y\_{i}-(a\_{0}+ a\_{1}x\_{i}^{1}+…+ a\_{k}x\_{i}^{k}))^{2} $

ניתן לבדוק ולראות (אם לא ראיתם, אז [קישור](http://mathworld.wolfram.com/LeastSquaresFittingPolynomial.html)) שהמקדמים שממזערים את הפונקציה הזאת הם:

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

כאשר $y$ וקטור הערכים $y\_{i}$, $X$ מטריצת ונדרמונדה (עוד [קישור](http://mathworld.wolfram.com/VandermondeMatrix.html) 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 אותן.*

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

 $loss\left(a\right)=\frac{1}{2}\sum\_{i=1}^{n}(y\_{i}-(a\_{0}+ a\_{1}x\_{i}^{1}+…+ a\_{k}x\_{i}^{k}))^{2} +\frac{1}{2}λ a^{T}a$

*ניתן להוכיח שהפעם הנוסחה ל-*$a$ *נתונה על ידי:*

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

*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 האילוץ שלנו על הפרמטרים מחמיר מדי ולא מאפשר לנו ללמוד.*

1. *כמו שאמרנו בתרגול, כדי לקבוע שמימד* VC *של קבוצת פונקציות* $H$ *הוא* $d$*, צריך למצוא קבוצת נקודות בגודל* $d$ *ש-*$H$ *מנפצת ולהוכיח ש-*$H$ *לא מנפצת אף קבוצה בגודל* $d+1$*.*
2. *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-* $H\_{C}$ *: אפשר לקחת למשל אינטרוולים* $\left(0, 3\right), \left(0.5, 1.5\right), \left(0, 2\right), (1, 3)$ *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-*$H\_{C}$ *כזאת ש-* $h\left(c\_{1}\right)=h(c\_{3})=1$ *ואילו* $h\left(c\_{2}\right)=0$*, לכן* $H\_{C}$ *לא מנפצת אף קבוצה עם 3 איברים.*

1. *לכל קבוצה* $C$ *מתקיים כי* $\left|H\_{C}\right|\leq |H|$*. לכן, אם* $\left|H\right|<2^{\left|C\right|}$*, אז* $H$ *לא תנפץ את* $C$*. מכאן ש:* $VCdim\left(H\right)\leq log\_{2}|H|$*.*
2. *נבחר לשם נוחות* $C=\{\left(0, 1\right), \left(-0.5, 0\right), (0.5, 0)\}$*. 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):*

**