מבוא למערכות לומדות
תרגיל 3

31 במאי 2017

1 שקילות של הגדרות Soft-SVM

נראה את שקילות ההגדרות:

$$(1)\ \min_{\mathbf{w}} \frac{\lambda}{2}\left\|\mathbf{w}\right\|^2 + \frac{1}{m}\sum_{i=1}^{m}\ell^{hinge}\left(y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle\right)$$
$$(2)\ \min_{\mathbf{w},\{\xi_i\}} \frac{\lambda}{2}\left\|\mathbf{w}\right\|^2 + \frac{1}{m}\sum_{i=1}^{m}\xi_i \text{ such that } \forall i, y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle \geq 1-\xi_i \text{ and } \xi_i \geq 0$$

כלומר נראה שפתרון ל־(1) הוא פתרון חוקי ואופטימלי ל־(2), ולהיפך. נניח ש־$\mathbf{w}$ הוא פתרון ל־(1). נזכור כי $\ell^{hinge}\left(y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle\right) = \max\left\{0, 1-y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle\right\}$. יהיו $\{\xi_i\}$ כך שלכל i מתקיים $y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle \geq 1-\xi_i$ וגם $\xi_i \geq 0$. אם כך:

$$y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle \geq 1-\xi_i$$
$$\iff \xi_i \geq 1-y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle$$

כמוכן, $\xi_i \geq 0$ לכן: $\xi_i \geq \max\left\{0, 1-y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle\right\} = \ell^{hinge}\left(y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle\right)$. זה נכון לכל i וכל הביטויים הנדונים הם אי־שליליים. לכן:

$$\frac{1}{m}\sum_{i=1}^{m}\xi_i \geq \frac{1}{m}\sum_{i=1}^{m}\ell^{hinge}\left(y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle\right)$$

אם כך, נקבע $\xi_i' = \ell^{hinge}\left(y_i\left\langle \mathbf{w},\mathbf{x}_i\right\rangle\right)$ ונקבל ש־$\mathbf{w},\{\xi_i'\}$ הוא פתרון חוקי ואופטימלי ל־(2), כנדרש.

בכיוון השני, נניח ש־$\mathbf{w},\{\xi_i\}$ פתרון ל־(2).

2 גרעינים (Kernels)

נגדיר את המיפוי $\psi:\{M,\ldots,N\}\to\{0,1\}^N$ באופן הבא:

$$x \mapsto \left(\overbrace{11\cdots1}^{x}\overbrace{0\cdots0}^{N-x}\right)$$

לכל $x\in\{M,\ldots,N\}$. נראה שלכל $x,x'\in\{M.\ldots,N\}$ מתקיים $K\left(x,x'\right)=\left\langle\psi\left(x\right),\psi\left(x'\right)\right\rangle$. יהיו x,x' כנ"ל. נניח בה"כ ש־$x\leq x'$. לכן $M\leq\min\left\{x,x'\right\}=x\leq x'\leq N$. מתקיים:

$$\psi\left(x\right)=\left(\overbrace{11\cdots1}^{x}\overbrace{0\cdots0}^{N-x}\right)$$
$$\implies \forall i\in[N]\ \left(\psi\left(x\right)\right)_i \overset{*}{=} \begin{cases}1, & i\leq x\left(\leq x'\right)\\ 0, & \text{otherwise}\end{cases}$$

ובאופן דומה:

$$
\begin{aligned}
\left(\psi\left(x'\right)\right)_i &\overset{**}{=} \begin{cases} 1, & i \le x' \\ 0, & \text{otherwise} \end{cases} \\
\implies \left\langle \psi\left(x\right), \psi\left(x'\right) \right\rangle &\overset{1}{=} \sum_{i=1}^{N} \left(\psi\left(x\right)\right)_i \cdot \left(\psi\left(x'\right)\right)_i \\
&\overset{2}{=} \sum_{i=1}^{x} \overbrace{1 \cdot 1}^{i \le x \le x'} + \sum_{i=x+1}^{x'} \overbrace{1 \cdot 0}^{x < i \le x'} + \sum_{i=x'+1}^{N} \overbrace{0 \cdot 0}^{i > x'} \\
&= x \cdot 1 + \left(x' - x\right) \cdot 0 + \left(N - x'\right) \cdot 0 = x \\
&= \min\left\{x, x'\right\} = K\left(x, x'\right)
\end{aligned}
$$

כאשר:

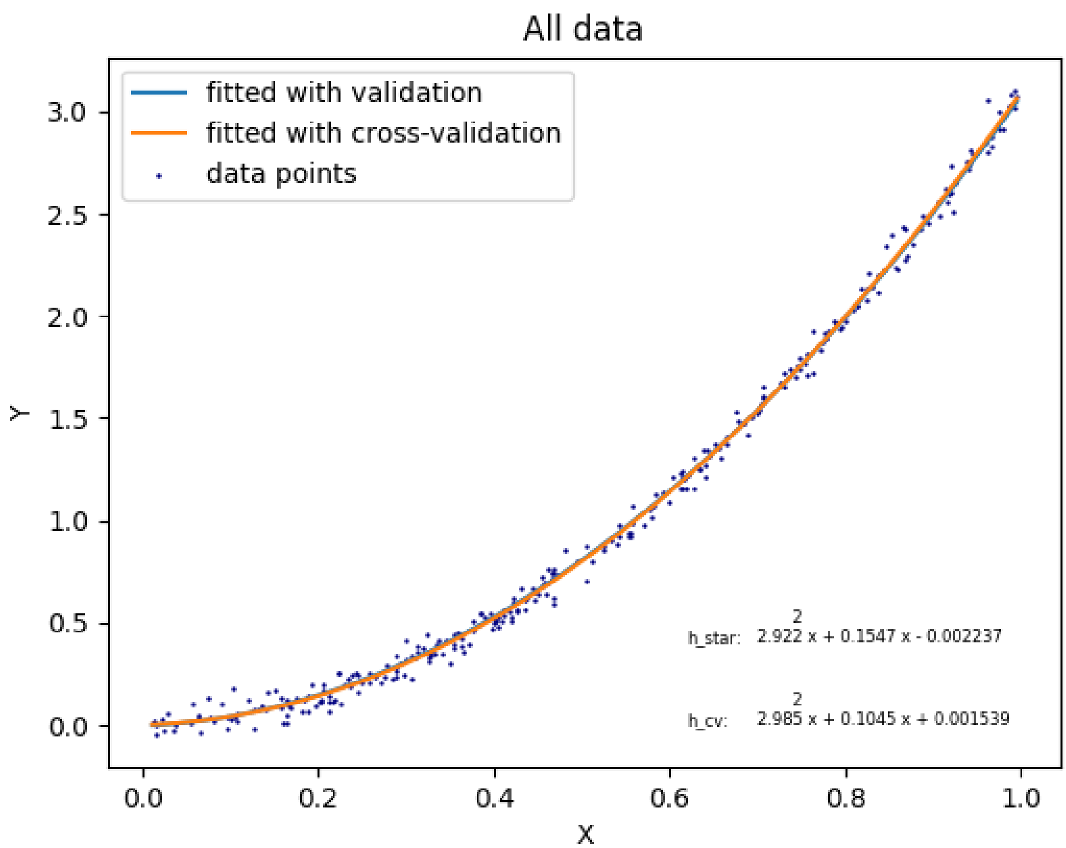
1. לפי הגדרת המכפלה הפנימית
2. נובע מ־$(*)$ ו־$(**)$. נשים לב שהסכום האמצעי יכול להיות ריק (אם $x = x'$)

3 בחירת מודל

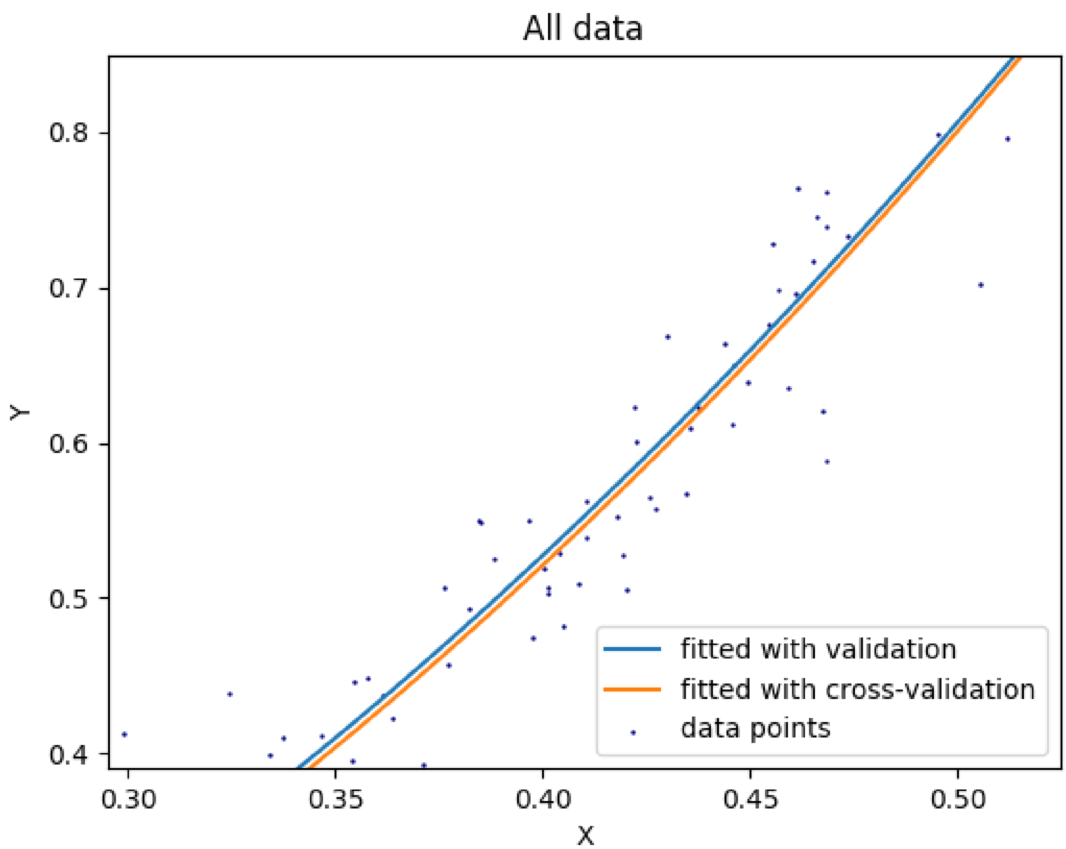
4 חלק תכנותי - Validation

שגיאת ההיפותזה h^* שהתקבלה בתהליך על נתוני ה־test היא: 0.02.

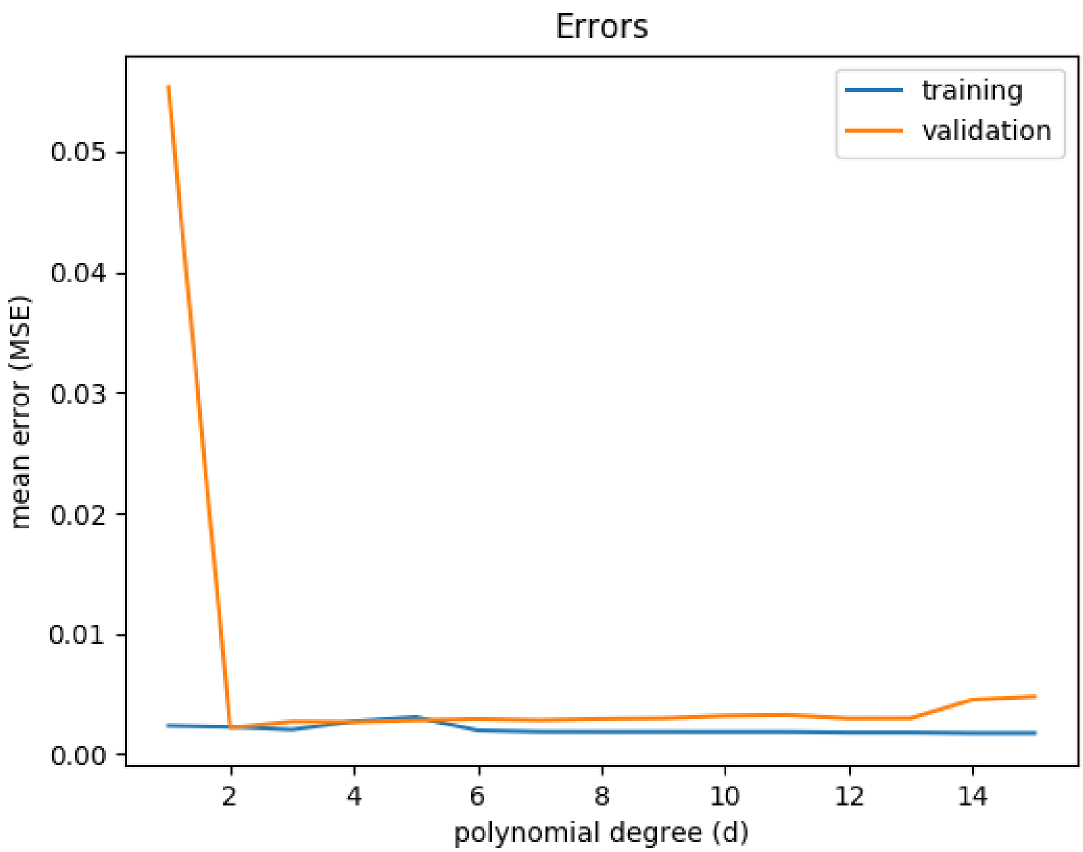
ביצוע k-fold cross validation מחזיר היפותזה דומה ל־h^*, אבל לא זהה (הפרשים קטנים במקדים).



איור 1: אפשר לראות שהפולינום h^* שהותאם בתהליך ה־validation כמעט זהה ל־h_{cv} שהותאם ב־cross-validation



איור 2: קלוז־אפ: המרחק בין h^* ל־h_{cv} זניח



איור 3: ככל שדרגת הפולינום גדלה, שגיאת האימון קטנה אך שגיאת הולדיציה עולה קצת (בגלל overfitting). עבור פולינום מדרגה 1 ברור שיש underfitting ולכן שגיאת הולידציה גדולה.