IML תרגיל 3

בר אלוני [redacted]

Bayes Optimal and LDA

1. עבור $x \in X$ כלשהוא, נתבונן ב: $argmax_{y\in\{\pm1\}} \Pr(x|y) \Pr(y)$ כפונקציה מ-X ל-$\{\pm1\}$:

$$argmax_{y\in\{\pm1\}} \Pr(x|y)\Pr(y) = argmax_{y\in\{\pm1\}} \frac{\Pr(y|x)}{\Pr(x)}$$

ומכיוון שהמונה בלבד תלוי ב-y, אז ה-y שממקסם את המונה ממקסם גם את הביטוי כולו ולכן למעשה הביטוי מעלה שקול ל: $argmax_{y\in\{\pm1\}} \Pr(y|x)$.

נבחין ש:

$$\mathrm{h_D}(\mathrm{x}) = 1 \Leftrightarrow \Pr(y=1|x) \geq \frac{1}{2} \Leftrightarrow \Pr(y=1|x) \geq \Pr(y=-1|x) \Leftrightarrow argmax_{y\in\{\pm1\}} \Pr(y|x) = 1$$

כאשר המעבר השני נובע מכן ש-y יכול לקבל רק את הערכים {1, -1}, כלומר המאורעות המוצגים משלימים. נשים לב שמהביטוי מעל גם משתמע ש:

$$\mathrm{h_D}(\mathrm{x}) = -1 \Leftrightarrow argmax_{y\in\{\pm1\}} \Pr(y|x) = -1$$

כנדרש. ■

2. מתקיים מסעיף קודם:

$$h_D(x) = argmax_{y\in\{\pm1\}} f(x|y)\Pr(\mathrm{y})$$

נציב ונקבל:

$$argmax_{y\in\{\pm1\}} \left(\frac{1}{((2\pi)^d \det(\Sigma))^{\frac{1}{2}}}\right) \exp\left\{-\frac{1}{2}(x-\mu_y)^T \Sigma^{-1}(x-\mu_y)\right\} \Pr(y)$$

מכיוון שפונקציית הלוג היא מונוטונית עולה, אז מקסום הביטוי הנ"ל שקול למיקסום לוג הביטוי. כלומר הביטוי מעל שקול לביטוי הבא:

$$argmax_{y\in\{\pm1\}} \ln\left(\left(\frac{1}{((2\pi)^d \, det(\Sigma))^{\frac{1}{2}}}\right) exp\left\{-\frac{1}{2}(x-\mu_y)^T \Sigma^{-1}(x-\mu_y)\right\} Pr(y)\right) =$$

$$argmax_{y\in\{\pm1\}} \, ln\left(\frac{1}{((2\pi)^d \det(\Sigma))^{\frac{1}{2}}}\right) - \frac{1}{2}(x-\mu_y)^T \Sigma^{-1}(x-\mu_y) + \ln(Pr(y)) =$$

$$argmax_{y\in\{\pm1\}} \, ln\left(\frac{1}{((2\pi)^d \det(\Sigma))^{\frac{1}{2}}}\right) - \frac{1}{2}\left(x^T\Sigma^{-1}x - \mu_y^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_y + \mu_y^T\Sigma^{-1}\mu_y\right) + \ln(Pr(y))$$

נתעלם מהביטויים הצבועים בכחול, שאינם תלויים ב- y (מכיוון שאם נמקסם את הביטוי ללא הקבועים, נמקסם כמובן את הביטוי עם הקבועים) ונקבל:

$$argmax_{y\in\{\pm1\}} \, x^T\Sigma^{-1}\mu_y - \frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y + \ln(Pr(y))$$

כאשר המעבר האחרון נובע מכך ש: x, μ_y קבועים ולכן: $\mu_y^T\Sigma^{-1}x = x^T\Sigma^{-1}\mu_y$, וקיבלנו כנדרש.

■

Spam

3. 2 סוגי הטעויות שהמסווג שלנו יכול לעשות הם:
   1. לסווג דואר זבל כדואר רגיל.
   2. לסווג דואר רגיל כדואר זבל.

נבחין שהשגיאה השנייה חמורה יותר, לא נרצה לפספס הודעה חשובה מכיוון שסיווגנו אותה כזבל בטעות. מכיוון שהמוסכמה היא שהשגיאה החמורה יותר תהיה type I error (כלומר סימנו דגימה כחיובית: 1, כשהיא למעשה שלילית: -1), נצטרך לתייג כך: $\{spam, not-spam\} = \{1, -1\}$ כדי ששגיאה 2 אותה תארנו אכן תהיה type I error.

SVM- Formulation

4. ראשית נזכור ש: $\|w\|^2 = \langle w, w\rangle = w^T w = w^T I_d w$.
כלומר נצטרך להגדיר את הפרמטרים של בעיית ה-QP המתאימה כך:
   - $Q = 2I_d$
   - $a = 0 \in \mathbb{R}^d$
   - $v = w$
   - $A = \begin{bmatrix} & \vdots & \\ \vdash & -y_i x_i^T & \dashv \\ & \vdots & \end{bmatrix}$
   - $d = y_i b - 1 \in \mathbb{R}^d$

ואכן נקבל שבעיית ה:hard-SVM היא מהצורה המתבקשת:
$$\frac{1}{2} v^T Q v + a^T v = \frac{1}{2} * 2w^T I_d w + \langle 0, w\rangle = w^T I_d w = \|w\|^2$$
בעלת אילוצים שקולים:

$$Av \leq y_i b - 1 \in \mathbb{R}^d \Leftrightarrow \begin{bmatrix} & \vdots & \\ \vdash & -y_i x_i^T & \dashv \\ & \vdots & \end{bmatrix} w \leq y_i b - 1 \in \mathbb{R}^d \Leftrightarrow$$
$$\forall i:\ -y_i x_i^T w \leq y_i b - 1 \in \mathbb{R}^d \Leftrightarrow \forall i:\ -y_i \langle x_i, w\rangle \leq y_i b - 1 \in \mathbb{R}^d \Leftrightarrow$$
$$\forall i: -y_i \langle w, x_i\rangle \leq y_i b - 1 \in \mathbb{R}^d \Leftrightarrow\ \forall i: y_i \langle w, x_i\rangle \geq 1 - y_i b \in \mathbb{R}^d \Leftrightarrow$$
$$\forall i:\ y_i \langle w, x_i\rangle + y_i b \geq\ 1 \in \mathbb{R}^d \Leftrightarrow \forall i:\ y_i(\langle w, x_i\rangle + b) \geq\ 1 \in \mathbb{R}^d$$
■

5. האילוצים בבעית ה- soft-SVM הם: $\forall i: \xi_i \geq 0\ and\ y_i\langle w, x_i\rangle \geq 1 - \xi_i$
כלומר בביטוי (1) אנחנו שמים בסכום משתנה שמקיים את שני התנאים הללו.
נבחן את הפונקציה שאנו שמים בסכום בביטוי (2):
$\ell^{hinge}(1 - y_i\langle w, x_i\rangle) = \max\{0, (1 - y_i\langle w, x_i\rangle)\}$.

מתקיים:
   - $\ell^{hinge}(1 - y_i\langle w, x_i\rangle) = 0 \Leftrightarrow 1 - y_i\langle w, x_i\rangle \leq 0 \Leftrightarrow y_i\langle w, x_i\rangle \geq 1 \Leftrightarrow$
   $y_i\langle w, x_i\rangle \geq 1 - \xi_i\ for\ \xi_i \geq 0$
   - $\ell^{hinge}(1 - y_i\langle w, x_i\rangle) = 1 - y_i\langle w, x_i\rangle \Leftrightarrow 1 - y_i\langle w, x_i\rangle > 0 \Leftrightarrow$
   $1 > y_i\langle w, x_i\rangle \Leftrightarrow y_i\langle w, x_i\rangle \geq 1 - \xi_i\ for\ \xi_i \geq 0$

   כאשר המעבר האחרון נעשה ע"י שימוש בתכונת הסופרימום.

כלומר מצאנו שהביטויים בסכומים ב-2 ההגדרות של soft-SVM מסכימות, ולכן ההגדרות שקולות, כנדרש.

■

Implemention and simulation-comparison of different classifiers

6. בקובץ
7. בקובץ