למידת מכונה 5

14 במאי 2020

1 תרגיל 1:

הגדרנו את $h_{\mathcal{D}}(x)=\begin{cases}1 & \mathcal{P}(y=1|x)\geq\frac{1}{2}\\ -1 & otherwise\end{cases}$. ניזכר כי לפי כלל בייס $\mathcal{P}(x|y)\mathcal{P}(y)=\mathcal{P}(y|x)\mathcal{P}(x)$ מצד שני מאחר ו $\sum\mathcal{P}(y|x)=1$ ו- $y\in\{-1,1\}$ אז מתקיים כי אם $\mathcal{P}(y=1|x)\geq\frac{1}{2}$ אז- $\mathcal{P}(y=1|x)=\max_{y\in\{-1,1\}}\{\mathcal{P}(y|x)\}$. באותו אופן עבור עבור $\mathcal{P}(y=-1|x)\geq\frac{1}{2}$.
ולכן:

$$\begin{aligned}h_{\mathcal{D}}(x)&=\arg\max_{y\in\{-1,1\}}\mathcal{P}(y|x)=\mathcal{P}(x)\arg\max_{y\in\{-1,1\}}\mathcal{P}(y|x)\overset{(1)}{=}\\&=\arg\max_{y\in\{-1,1\}}\mathcal{P}(y|x)\mathcal{P}(x)=\arg\max_{y\in\{-1,1\}}\mathcal{P}(x|y)\mathcal{P}(y)\end{aligned}$$

כאשר הכנסנו את x מאחר והוא מגדיל את כל האיברים מהם אנו מחלצים את המקסימום באופן זהה.

2 תרגיל 2:

ממונטוניות חזקה של פונקציית ה $\ln$ נקבל כי $\arg\max_y g(y)=\arg\max_y\ln(g(y))$. ולכן

$$\begin{aligned}\Rightarrow h_{\mathcal{D}}(x)&=\arg\max_{y\in\{-1,1\}}\mathcal{P}(x|y)\mathcal{P}(y)=\arg\max_{y\in\{-1,1\}}\ln(\mathcal{P}(x|y)\mathcal{P}(y))=\\&=\arg\max_{y\in\{-1,1\}}\left\{\ln\left(\frac{1}{\sqrt{(2\pi)^d\det(\Sigma)}}e^{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)}\right)+\ln(\mathcal{P}(y))\right\}=\\&=\arg\max_{y\in\{-1,1\}}\left\{\underbrace{\ln\left(\frac{1}{\sqrt{(2\pi)^d\det(\Sigma)}}\right)}_{not\,depend\,at\,y}+\ln\left(e^{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)}\right)+\ln(\mathcal{P}(y))\right\}=\\&=\arg\max_{y\in\{-1,1\}}\left\{\ln\left(e^{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)}\right)+\ln(\mathcal{P}(y))\right\}=\arg\max_{y\in\{-1,1\}}\left\{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)+\ln(\mathcal{P}(y))\right\}=\\&\arg\max_{y\in\{-1,1\}}\left\{\underbrace{-\frac{1}{2}x^T\Sigma^{-1}x}_{not\,depend\,at\,y}+x^T\Sigma^{-1}\mu_y-\frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y+\ln(\mathcal{P}(y))\right\}=\\&\arg\max_{y\in\{-1,1\}}\left\{x^T\Sigma^{-1}\mu_y-\frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y+\ln(\mathcal{P}(y))\right\}=\arg\max_{y\in\{-1,1\}}\delta_y\end{aligned}$$

3 תרגיל 3:

את Σ ו $\mathcal{P}$ נחשב על ידי החישוב הרגיל על ה$training-samples$. (חישוב ישיר) מה שמשתנה ולא היה בשאלות עד כה זה החישוב של μ_+ ו μ_- אבל גם כאן זה מאוד פשוט. $\mu_\pm=Mean\{x|(x,y)=(x,\pm1)\}$.

4 תרגיל 4:

נירצה שהבחינה של הודעה נורמטיבת כהודעת ספאם תיהיה: $Type-2-error$. ולכן נתייג כ $False\,Positive.=\{not-spam,true\}$ ואת: $True\,Positive.=\{spam,\,false\}$.

5 תרגיל 5:

אני מצרף תמונה בכתב יד מאחר וזה הרבה יותר ברור ככה :

SVM - Formulation 5

נבחר את $Q = 2I$ ואת $a = \underline{0}$ ונקבל כי :

$$\forall_j \quad y_j \cancel{sign} \cdot (\langle w, x_j \rangle + b) \geq 1$$

$$\iff \quad y_j \langle w, x_j \rangle \geq 1 - b y_j$$

$$\Rightarrow \begin{bmatrix} - y_1 - \\ - y_2 - \\ \vdots \\ - y_m - \end{bmatrix} \begin{bmatrix} - w - \\ - w - \\ \vdots \\ - w - \end{bmatrix} \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix} \geq \begin{bmatrix} 1 - b y_1 \\ \vdots \\ 1 - b y_m \end{bmatrix}$$

$$\Rightarrow \underbrace{- \begin{bmatrix} - y_1 - \\ \vdots \\ - y_j - \end{bmatrix} \begin{bmatrix} - w - \\ \vdots \\ - w - \end{bmatrix}}_{A = -y^T \mathbb{1}_{m \times m} w} \underbrace{\begin{bmatrix} | & & | \\ x_1 & \cdots & x_m \\ | & & | \end{bmatrix}}_{v} \leq \underbrace{\begin{bmatrix} 1 - b y_1 \\ \vdots \\ 1 - b y_m \end{bmatrix}}_{d}$$

מטריצה של 1-ים

6 תרגיל 6:

נראה שקילות בין שתי הבעיות :

$$\min_{\omega,\xi_i,y_i\langle\omega,x_i\rangle\geq1-\xi_i} \frac{\lambda}{2}||\omega||^2+\frac{1}{m}\sum_{i=1}^{m}\xi_i = \min_{\omega,\omega,\xi_i,y_i\langle\omega,x_i\rangle\geq1-\xi_i}\left\{\frac{\lambda}{2}||\omega||^2+\frac{1}{m}\sum_{i=1}^{m}\min_{\xi_i}\xi_i\right\} =$$

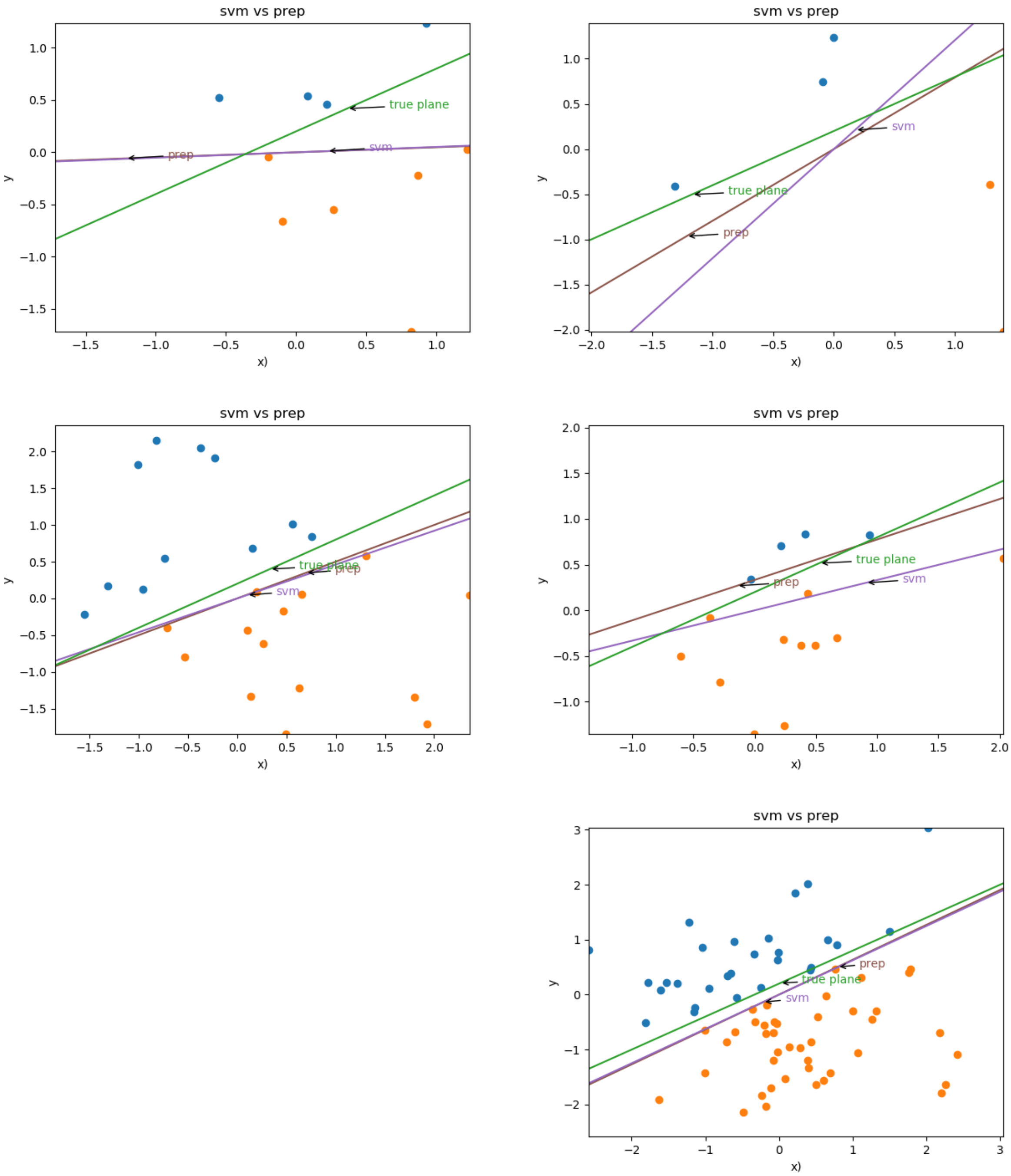
$$= \min\left\{\frac{\lambda}{2}||\omega||^2+\frac{1}{m}\sum_{i=1}^{m}\begin{cases}1-y_i\langle\omega,x_i\rangle & if\ 1-y_i\langle\omega,x_i\rangle\geq0\\ 0 & otherwise\end{cases}\right\} =$$

$$= \min_{\omega}\left\{\frac{\lambda}{2}||\omega||^2+\frac{1}{m}\sum_{i=1}^{m}\max\left\{0,1-y_i\langle\omega,x_i\rangle\right\}\right\}$$

נשים לב שכל מעבר הוא בשיוון ולכן הבכרח פתרון של בעיה אחת פותר גם את הבעיה השניה.

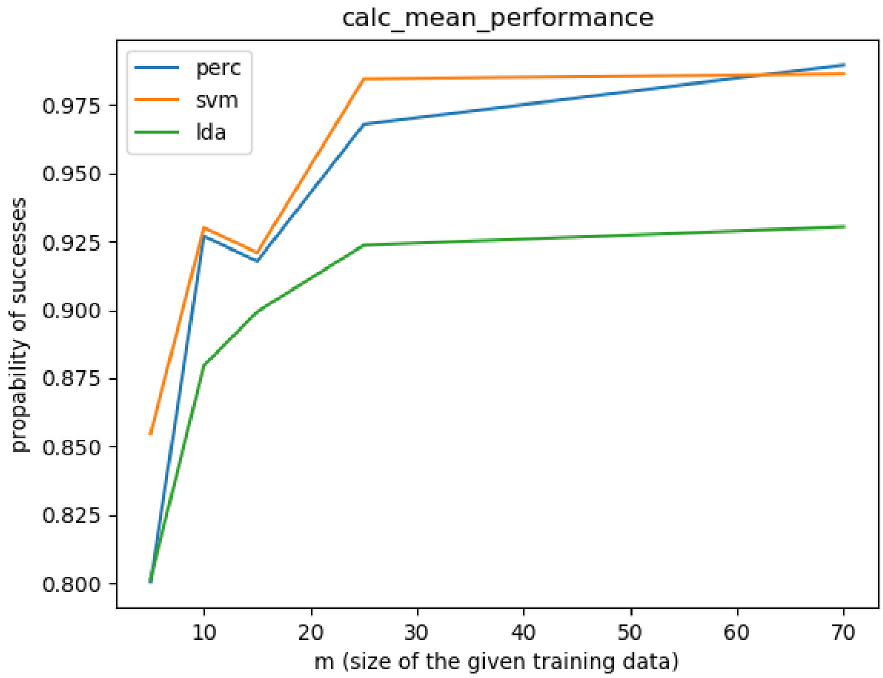
7 השוואה בין שיטות למציאת ישר מפריד.

כל השיטות נמצאו יעילות מבחינת יכולת הדיוק, נציין ש LDA איטית באופן משמעותי, בעיקר השלב של מציאת המטריצה ההופכית ($\Sigma^{\dagger}$). הגרפים של $preception$ לעומת svm . ניתן ממש ליראות איך המישורים מתאחדים :



7.1 ההשוואה בין בשיטות השונות לLDA.

קל לראות ש LDA לא נותן לנו שיערוך טוב. (כלומר כן טוב אבל פחות) אני מאמין שהוא סובל יותר מאובר פיט. מאחר ומראש הוא לא בהכרח מייצג מישור ולכן ספיציפית לבעייה הזאת הוא פחות מתאים.



8 זיהוי ספרות :

ההשוואה על זיהוי הספרות, אני חייב להגיש שממש הואשמתי מי מידת הדיוק. שיטת ה SVM לקחה את רוב זמן החישוב כאשר על הקלטים הגדולים זמן הריצה היה באזור $\sim 0.05\ [sec]$. כל שאר השיטות ירדו מ 0.025 שניות עבור הקלטים הגדולים.