זיהוי תבניות ולמידה חישובית.

תרגיל בית מספר 3

1. (15) בבעיית סווג של שתי מחלקות עם וקטור תכונות חד-מימדי x, ידוע כי ההתפלגויות האפריוריות הן שוות, כלומר $p(\omega_1) = p(\omega_2)$. מהו כלל ההחלטה אם ידוע כי:

$$p(x|\omega_i) = \begin{cases} \dfrac{x}{\sigma_i^2} e^{\left(-\frac{x^2}{2\sigma_i^2}\right)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

2. (15) בבעיית סווג לשתי מחלקות, מדדו תכונה אחת x (בעייה חד-מימדית). נתון כי פונקציות הצפיפות של התכונה x הן גאוסיות, כאשר פונקציית הצפיפות המותנית של התכונה עבור המחלקה הראשונה היא גאוסית עם ממוצע 1 ושונות $\sigma_1^{\;2}$,

$$f(x|\omega_1) = \frac{1}{\sqrt{2\pi\sigma_1^{\;2}}} e^{\frac{-(x-1)^2}{2\sigma_1^{\;2}}}$$

ופונקציית הצפיפות המותנית השניה היא גאוסית עם ממוצע 0 ושונות $\sigma_2^{\;2}$,

$$f(x|\omega_2) = \frac{1}{\sqrt{2\pi\sigma_2^{\;2}}} e^{\frac{-x^2}{2\sigma_2^{\;2}}}$$

כמו-כן ידוע כי מטריצת העלות היא המטריצה הבאה:

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} 0 & \lambda_{12} \\ \lambda_{21} & 0 \end{pmatrix}$$

חשבו את ערך הסף x0 למיזעור שגיאת הסיכון.

3. ממשו מסווג SVM ב- Matlab, באופן דומה לזה שמימשנו עבור הפרספטרון בתרגיל כיתה.

כתבו את הפונקציות הבאות: svm_train(X,y) ו- svm_test(theta,X_test,y_test).

השתמשו בפונקציה quadprog הפותרת בעיות אופטימיזציה קוודרטיות: quadprog(H,f,A,b)
מהצורה: $\frac{1}{2}x^T Hx + f^T x$ תחת האילוץ $Ax \le b$

<u>הסבר על הפונקציה quadprog</u>

הפונקציה quadprog פותרת בעיות אופטימיזציה עם אילוצים (ראו ב- Matlab help).

`x = quadprog(H,f,A,b)` returns a vector `x` that minimizes `1/2*x'*H*x + f'*x` subject to `A*x ≤ b`.

או:

`x = quadprog(H,f,A,b,Aeq,beq,lb,ub,x0,options)` minimizes with the optimization options specified in the structure `options`. Use `optimset` to set these options. If you do not wish to give an initial point, set `x0 = []`.

מהם פרמטרי הכניסה (Input arguments) של הפונקציה? באופן ספיציפי, מה צריכים להיות הערכים של הפונקציות H,f,A,b כדי שנוכל לקבל את ערכי הוקטור θ של השוליים הגאומטריים המקסימליים?

(נבחין כי המשתנה x המופיע עבור הפונקציה quadprog עבור בעיית ה- SVM הוא לא וקטורי התכונות או מטריצת התכונות X, אלא וקטור המשקלות θ).

ראשית נתבונן בבעיית האופטימיזציה: עלינו למצוא מהו θ שיביא למינימום את הביטוי $\frac{1}{2}\|\theta\|^2$,

תוך עמידה בתנאי האילוץ.

בפונקציה quadprog הבעייה מנוסחת באופן הבא: $\frac{1}{2}x^T Hx + f^T x$. מהם H ו- f?

אפשר גם לרשום את הביטוי הקודם (אותו צריך להביא למינימום) באופן הבא:

$$\frac{1}{2}\|\theta\|^2 = \frac{1}{2}(\theta^T \cdot \theta) = \frac{1}{2}(\theta^T \cdot I \cdot \theta)$$

כאשר θ הוא הוקטור הניצב לעל-מישור המפריד המבוקש, ו- I היא מטריצת היחידה:

$$\|\theta\|^2=\theta^T\theta=(\theta_1\ \theta_2\ \ldots\theta_d)\cdot\begin{pmatrix}\theta_1\\ \theta_2\\ \\ :\\ \theta_d\end{pmatrix}=(\theta_1\ \theta_2\ \ldots\theta_d)\cdot\begin{pmatrix}1&0&..&.&0\\0&1&..&.&0\\0&0&&&0\\:&:&&:&0\\0&0&&0&1\end{pmatrix}\begin{pmatrix}\theta_1\\ \theta_2\\ \\ :\\ \theta_d\end{pmatrix}=\theta^T\cdot I\cdot\theta$$

כלומר הפונקציה H היא מטריצת היחידה. בבעייה שלפנינו המימד הוא d=2, ולכן:

$$\|\theta\|^2=\theta^T\theta=(\theta_1\ \theta_2)\cdot\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}=(\theta_1\ \theta_2\ )\cdot\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}=\theta^T\cdot I\cdot\theta$$

כלומר $\min(\frac{1}{2}\|\theta\|^2)$ זהה לביצוע $\min(\theta_1\ \theta_2\ )\cdot\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}$, אבל הביטוי מימין במשוואה הקודמת הוא לפי הדרישה של הפונקציה quadprog.

לפיכך בפונקציה quadprog הערך של H צריך להיות מטריצת היחידה, כלומר $I_2=\begin{pmatrix}1&0\\0&1\end{pmatrix}$

או במונחים של Matlab:

```
H=[1 0; 0 1 ]
```

עתה נתבונן בתנאי האילוץ וננסה להבין כיצד לקודד אותו עבור הפונקציה quadprog.

תנאי האילוץ הוא $y_t\cdot\theta^T x_t\geq 1\quad for\ all\ t=1,2...,n$. נרצה לקודד אותו כפי שמופיע תנאי האילוץ עבור quadprog: `A*x ≤ b`

כאשר A היא מטריצת המקדמים המבטאת את המקדמים של רכיבי וקטור התכונות x באי השוויון, ו-b הם אגף ימין של כל אחד מאי-השוויונים.

לכן:

$$y_t\cdot\theta^T x_t\geq 1\quad for\ all\ t=1,2...,n\qquad\triangleq\qquad\begin{pmatrix}y_1&0&0&..&0\\0&y_2&0&..&0\\0&0&&&0\\&&&&:\\0&0&0&..&y_n\end{pmatrix}\begin{pmatrix}x_1{}^1&x_2{}^1&x_3{}^1&..&x_d{}^1\\x_1{}^2&x_2{}^2&x_3{}^2&..&x_d{}^2\\x_1{}^3&x_2{}^3&x_3{}^3&&x_d{}^3\\:&&&&:\\x_1{}^n&x_2{}^n&x_3{}^n&..&x_d{}^n\end{pmatrix}\begin{pmatrix}\theta_1\\ \theta_2\\ \theta_3\\ :\\ \theta_d\end{pmatrix}\geq\begin{pmatrix}1\\1\\1\\:\\1\end{pmatrix}$$

כפי שאפשר לראות הכתיבה המטריצית מאפשרת לכלול את כל האי-שוויונים של האילוצים $y_t \cdot \theta^T x_t \geq 1 \quad for\ all\ t = 1,2...,n$ כלומר שכל דוגמאות האימון תהיינה מסווגות נכונה עם מרווח גדול או שווה 1 (הראו באופן מפורש כי ההצגה המטריצית אכן מבטאת כל כל אי-השוויונים של תנאי האילוץ).

מכאן שכדי לקודד את תנאי האילוץ, המטריצה A צריכה להיות המכפלה של מטריצה אלכסונית בה האלכסון הראשי הוא וקטור התגיות (labels) y, עם מטריצה X בה דוגמאות האימון הן שורות המטריצה (כלומר כל וקטור תכונות הוא שורה), ווקטור המשקלות θ הוא מייצג את הוקטור המופיע בתנאי האילוץ של quadprog כוקטור עמודה.

מאחר ועבור quadprog אי השוויון הוא A*x ≤ b ואילו אצלנו אי השוויון הוא $D_y \cdot X \cdot \theta \geq 1$

נקודד את המטריצות ואת אגף ימין עם סימן שלילי, כלומר:

```
theta=zeros(d,1);
H = eye(d);
f = zeros(d,1);
Dy=diag(y);
A =Dy*X;
b=ones(size(y));
```

כך שהקריאה לפונקציה quadprog היא:

```
[theta fval ex op l] = quadprog(H, f, -A, -b, [], [], [], [], [], ...
 optimset('LargeScale', 'off', 'MaxIter', 10000, 'Display', 'off'));
```

4. הפעילו את המסווג על שתי קבוצות הנתונים מתרגיל הפרספטרון. מה ההבדל בין ערכי theta שהתקבלו כאן לבין אלה שהתקבלו בתרגיל הפרספטרון (חזרו על תרגיל הפרספטרון וחשבו שוב את ערכי theta)? האם יש צורך לחשב את ההפרש בין הערכים? ציירו בקו מקווקו את השוליים משני צידי העל-מישור המפריד.

5. עבור גבולות ההחלטה שחושבו על-ידי ה- SVM, חשבו את השוליים הגאומטריים המתאימים. השוו בין שוליים אלה לשוליים הגאומטריים שחישבתם עבור הפרספטרון (חזרו על תרגיל הפרספטרון וחשבו את השוליים הגאומטריים).

6. עתה הורידו את מסווג ה- SVM הבא: libSVM מתוך האתר:
http://www.csie.ntu.edu.tw/~cjlin/libsvm/

והפעילו אותו על קבוצות הנתונים הקודמות ועל תמונות של ספרות בכתב יד המצורפות בקבצים. העזרו בהדרכה הבאה:
http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

קבצי התמונות הם של ארבעה סוגי ספרות "2", "3", "4" ו-"8". הספרה הראשונה בשם הקובץ מתארת את סוג הספרות שבו. התמונות הן מטריצות בגודל 15X15 בגווני אפור. דרך נוחה לצפות בתמונת הספרה היא באמצעות הפקודה imagesc או imshow.

מהי שגיאת האימון, כלומר מהו החלק של דוגמאות האימון שלא סווגו נכונה? בחרו אחת מדוגמאות האימון שלא סווגו נכון. מדוע לדעתכם ה- SVM נכשל בסווג? עתה הפעילו את ה-SVM על קבוצת המבחן (test set) מהי שגיאת המבחן כלומר מהו החלק של דוגמאות המבחן שלא סווגו נכונה?

7. בצעו אופטימיזציה למסווג על-ידי בחירות שונות של הפרמטרים C ו- γ. העזרו בהדרכה במאמר מסעיף 6. ציירו גרף של שגיאת האימון כפונקצייה של C עבור ערכי γ שונים. עבור איזה ערכים מתקבלת שגיאת האימון הקטנה ביותר? האם תוכלו לבצע אופטימיזציה כך שהערך של C ו- γ יביא לשגיאת מבחן מינימלית?

הערות והוראות להפעלת libSVM

1. לפני הפעלת האימון והסווג באמצעות libSVM צריך לנרמל את הנתונים.

קראו במאמר [A practical guide to SVM classification](#) בסעיף 2.2 וכן B.

מצ"ב הפונקציה לנירמול הנתונים:

```
function [trainInput_n testInput_n] = normalizeTrainAndTest(trainInput,
testInput)
% normalizeTrainAndTest normalizes train and test data
% Input arguments:
% trainInput - train  matrix of feature vectors (each column is a
feature
% vector.
% testInput - test matrix of feature vectors
% Output arguments:
% trainInput_n - train matrix of feature vectors after scaling
% testInput_n - test matrix of feature vectors after scaling
% Usage: [trainInput
testInput]=normalizeTrainAndTest(trainInput,testInput);
%

[r1 c1] = size(trainInput);
[r2 c2] = size(testInput);

min1=min(trainInput);
max1=max(trainInput);
save min1max1 min1 max1;

mmin1=repmat(min1',1,r1);
mmax1=repmat(max1',1,r1);
mmin2=repmat(min1',1,r2);
mmax2=repmat(max1',1,r2);
trainInput_n=(trainInput-mmin1')./(mmax1'-mmin1');
testInput_n=(testInput-mmin2')./(mmax2'-mmin2');
```

2. הפקודה לאימון ה- SVM היא:

```
s=0; % s svm_type : set type of SVM (default 0)

t=0; % for linear SVM, or use t=2 for a Gaussian kernel
                          ראו גם באתר של libSVM לאחר ה- Graphic interface
msvm = 800;     % memory
   עכשיו יש צורך לרשום את האפשרויות של ה- SVM. במקרה של SVM ליניארי אין
       צורך בערך של gamma. ערכי c שונים הם ערכי העלות של דוגמאות החורגות
                                                              מהשוליים.
svmoptions =  ['-s ' num2str(s) ' -t '  num2str(t) ' -c ',...
num2str(C(c))...
```

```
            ' -g ' num2str(gamma(g)) ' -m ' num2str(msvm)];
                   השלב הבא הוא אימון ה- SVM ליצירת המודל של ה- SVM.
model = svmtrain(training_labels_all, training_data_all, svmoptions);
        לאחר שאימנו את ה- SVM אפשר להפעיל אותו לסווג נתוני המבחן:
 [predicted_labels, accuracy, dec_values]
=svmpredict(testing_labels_all, testing_data_all, model);
```