

מטלת מנחה (ממ"ן) 12

הקורס: מבוא לראייה ממוחשבת - 22928

חומר הלימוד למטלה: יחידות 4-6

מס' השאלות: 3

סמסטר: 2016א

משקל המטלה: 5 נקודות

מועד אחרון להגשה: 11.12.2016

הערות:

- להגשה - קובץ ZIP הכולל:
 - דו"ח בפורמט PDF עם תשובות לכל שאלה בתוספת איורים כנדרש.
 - יש לתאר בדו"ח את כל תהליך העבודה עם הפרמטרים שהשתמשתם בהם.
 - בנוסף, יש לצרף את תוכניות ה-Python עם הערות בגוף התוכנית ועם קישור ברור לשאלה המתאימה.
 - הקוד לא בא להחליף את הדו"ח ואין להדביק קטעי קוד בדו"ח.

חלק 1 – MNIST:

בחלק הראשון של המטלה נתרגל אלגוריתמים לסיווג שהוצגו ביחידות 4-6, ע"י סיווג וניתוח של תמונות של ספרות מתוך MNIST dataset.
(לקריאה נוספת - https://en.wikipedia.org/wiki/MNIST_database).
הקובץ עם התמונות נמצא באתר, הוא מכיל 3 קבוצות של תמונות, אחת כוללת 50K תמונות, משמשת לאימון, השניה והשלישית מכילות 10K ומשמשות ל-test ול-validation. במטלה נשתמש בסט האימון והבדיקה בלבד.
הקריאה ל-dataset נעשית ע"י:

```
# Load the dataset
f = gzip.open('mnist.pkl.gz', 'rb')
train_set, valid_set, test_set = cPickle.load(f)
f.close()

# separate to data and labels:
x_train, y_train = train_set
x_test, y_test = test_set
```

סט האימון מכיל מטריצה בגודל 50,000X768 שכל שורה בה היא תמונה של סיפרה אחרת, ווקטור באורך 50,000 שמכיל את ה-label או הסיפרה שמופיעה באמת בתמונה המתאימה. סט הבדיקה מכיל מטריצות בעלות מבנה דומה עם 10K שורות.

כדי להציג את התמונה ה-100 מתוך סט האימון:

```
# show sample #100 from the train set
i = 100
im = x_train[i].reshape([28,28])
label = y_train[i]

plt.figure()
plt.imshow(im, cmap='gray', interpolation='nearest')
plt.title(label)
plt.show()
```

ה-dataset מכיל תמונות של 10 ספרות. כמה תמונות יש מכל סיפרה? הצג את 12 הספרות הראשונות מתוך ה-train-set ב-figure אחד (מחולק ל 3 שורות וארבע עמודות).

הערה – לבעלי מחשב חלש, אפשר להשתמש ב-train-set המכיל את 10K התמונות הראשונות מה-dataset המקורי.

שאלה 1 – KNN

הדרך הפשוטה ביותר לסיווג היא ע"פ הדוגמא (או הדוגמאות) הדומות, מה שניקרא - Nearest Neighbor Classifier.

בשאלה זו נבדוק את הביצועים של מסווג KNN כאשר K הוא מספר השכנים.

a) חשב את ביצועי מסווג KNN עבור $k=1...10$. צייר את התוצאות על גרף. מה התוצאה האופטימלית? איך מתבצעת ההחלטה כאשר k זוגי?

שאלה 2 – PCA

בשאלה זו נבדוק את השימוש של PCA לצורך סיווג תמונות של מספרים כתובים בכתב יד מתוך MNIST.

(b) חשב את טרנספורמצית PCA על סט האימון. צייר את הסיפורה הממוצעת ואת 6 ה-principle components הראשונים.

(c) צייר גרף של הווריאנס הכולל ש"מוסבר" ע"י n הרכיבים הראשיים.

(d) כמה בסיסים צריך בשביל להגיע ל-95% ווריאנס? כמה ל-80%?

(e) הטל את הספרות למימד 2 וצייר את הווקטורים המתקבלים (נקודות על המישור, scatter plot), כל סיפורה בצבע אחר.

(f) חזור על שאלה 1 כאשר כל סיפורה מיוצגת ע"י ההטלה שלה למימד 2, 10 ו-20.

(g) עבוד סיפורה כלשהי, הטל את הסיפורה למימד k ושחזר אותה בחזרה. צייר את השיחזור עבור $k=2, 5, 10, 50, 100, 150$. כתוב את הנוסחה שבה השתמשת להטלה ולשיחזור.

(h) הבעיה העיקרית בשימוש ב-PCA היא הקושי בתיאור של dataset מורכב, כמו במקרה של שלנו. הפיתרון הוא לחשב טרנספורמצית PCA לכל מחלקה/סיפורה בנפרד בעזרת סט האימון. ולמדוד את ההתאמה של סיפורה חדשה לכל אחד מהמודלים.

1. חשב מודל PCA לכל סיפורה בנפרד. הצג את 6 ה-principle components של כל מודל. מה ההבדל יחסית למודל של כל הספרות יחד?

2. חשב את ההטלה של התמונות מה-test set לכל אחד מהמודלים (סה"כ 10 הטלות לכל תמונה).

3. שחזר את התמונות מכל אחת מההטלות (סה"כ 10 תמונות משוחזרות מכל תמונה). צייר את כל השיחזורים של תמונה אחת לדוגמה.

4. חשב את המרחק בין השיחזורים לתמונה המקורית.

5. המודל ששיחזר "הכי טוב" יבחר כמודל הנכון.

מה הביצועים של מסווג כזה?

חלק 2 – סיווג ע"י BOW

בחלק זה של העבודה נממש מסווג מסוג SVM שסיווג תמונות לשתי מחלקות, כאשר התמונות מיוצגות ע"י Bag Of Words (BOW).

נשתמש בשירות —Panoramio, שמאפשר לקבל סטים של תמונות ע"פ מספר קריטריונים, וביניהם גם ע"פ מיקום (geo-tag).
לדוגמא כתיבת:

```
http://www.panoramio.com/map/get_panoramas.php?
order=popularity&set=public&from=0&to=20&minx=35.234883&miny=31.777106&maxx=35.236883&maxy=31.779106&size=original
```

בשורת הכתובות בדפדפן, תפתח קובץ PHP שמכיל כתובות של 20 תמונות שצולמו בסביבת הר הבית.

פרטים נוספים אפשר לקרוא ב- <http://www.panoramio.com/api/data/api.html>, וגם בספר *Programming computer vision with python* בעמודים 69-63 (אפשר להוריד אותו מ- <http://programmingcomputervision.com>)

הנתונים שלנו בשאלה זו יתבססו על שתי סטים של תמונות שנדגמו דרך Panoramio משני מקומות בעלי מאפיינים חזותיים שונים. לדוגמה, סט אחד יכלול תמונות מתוך איזור עירוני וסט שני תמונות שצולמו בטבע. יש לדגום לפחות 200 תמונות מכל נקודה.

בשאלה זו נשתמש בתמונות מתוך ה-dataset הזה:

http://people.csail.mit.edu/torralba/code/spatialenvelope/spatial_envelope_256x256_static_8outdoorcategories.zip
המכיל תמונות מ-8 קטגוריות שונות כגון - הרים, חופים, יער. השייכות של כל תמונה לקטגוריה המתאימה נתונה בשם הקובץ. יש לבחור שתי קטגוריות מתוך השמונה. המטרה היא לבנות מסווג שיידע לסווג תמונה לקטגוריה הנכונה, כשהתמונה מיוצגת ע"י BOW.
חלקו את הנתונים לנתוני אימון - train-set, ונתוני בדיקה test-set ביחס של 4:1, כלומר 80% מהתמונות ישמשו לאימון ו-20% לבדיקה.

שלב האימון - יתבצע על ה- train-set בלבד:

1. חישוב מאפייני dense-SIFT מכל התמונות. (אפשר להשתמש בפונקציה **extract_features** מתוך OpenCV **לחישוב SIFT בנקודות מוגדרות מראש**).
2. Vector Quantization - חישוב K-means על סט המאפיינים. מספר המרכזים צריך להיות גדול מספיק (לפחות 100) ונתון לשיקולכם.
3. לכל תמונה חישוב היסטוגרמה של מספר המאפיינים השייכים ל-cluster ה-k.
4. אימון מסווג מסוג Support Vector Machine (SVM) לינארי כאשר כל תמונה מיוצגת ע"י ההיסטוגרמה שחושבה ב-(3) ולכל תמונה מוצמד label שמשייך אותה למחלקה המתאימה.

כתוצאה משלב האימון, המודל מורכב מהמרכזים שחושבו בשלב ה-VQ ע"י K-means והמסווג שאומן בשלב (4).

שלב הבדיקה - יתבצע על ה- test-set בלבד לכל תמונה -

1. חישוב מאפייני SIFT.
2. Vector Quantization - לכל מאפיין חישוב המרכז הקרוב ביותר.
3. חישוב היסטוגרמה של מספר המאפיינים בכל מרכז.
4. סיווג ע"י המסווג שאומן בשלב האימון.

מה הפרמטרים האופטימליים של המסווג. יש להציג את הביצועים עבור מספר ערכים של הקבוע C ע"י עקומות ROC וחישוב ה-AUC. יש להציג את הביצועים של המסווג ע"י ROC ו-AUC. איך יישתנו הביצועים עבור חלוקות שונות ל- train ו- test ? מה התוצאה הממוצעת?

בהצלחה,
אמיר