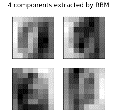
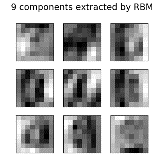
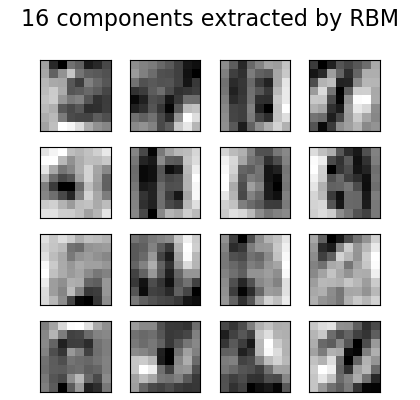
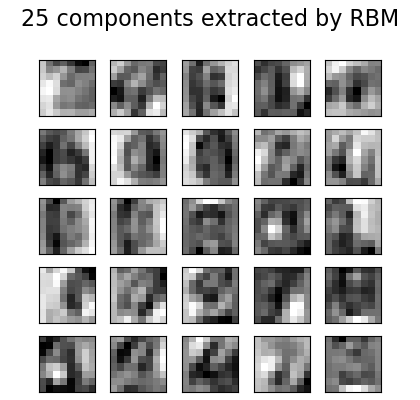
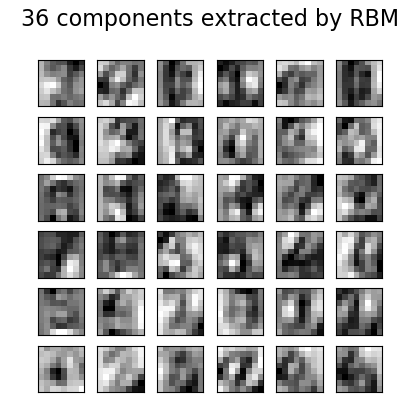
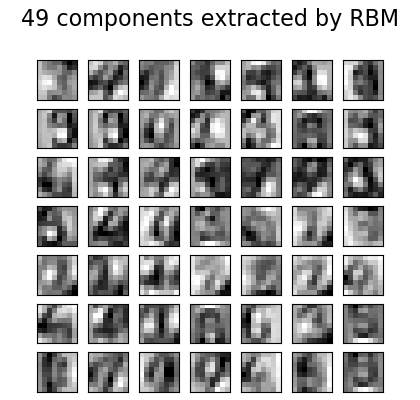
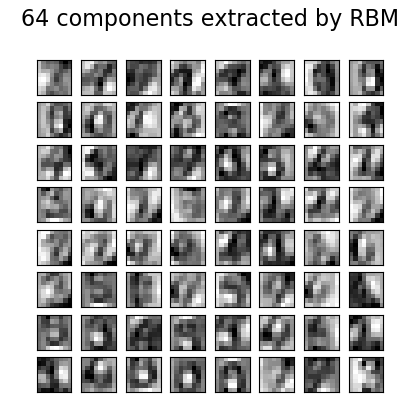
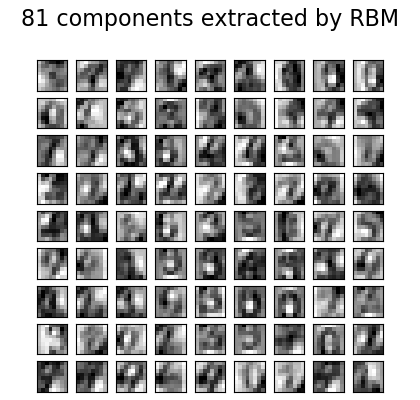
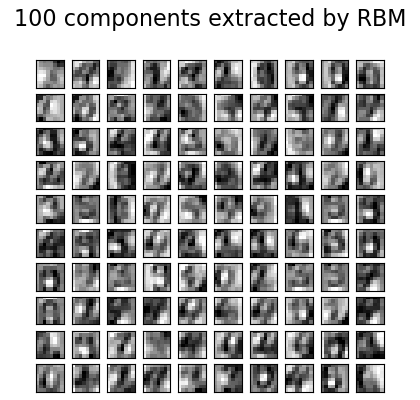
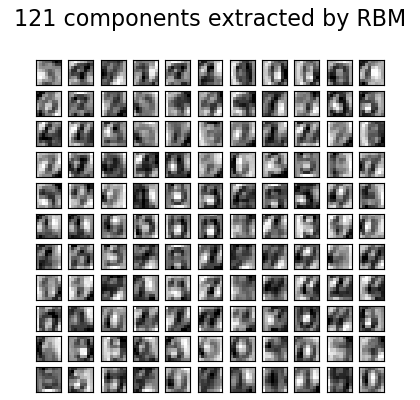
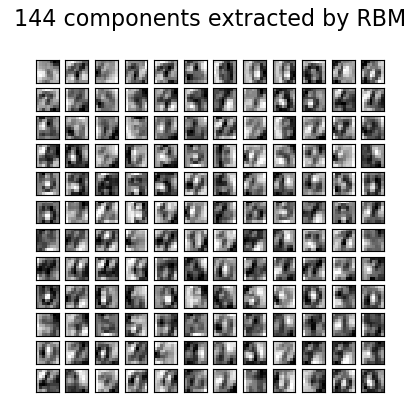
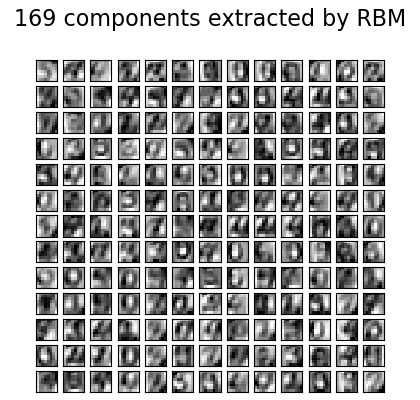
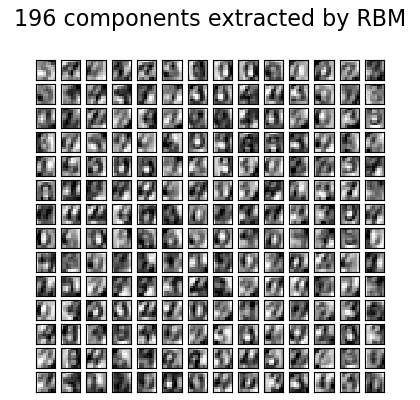
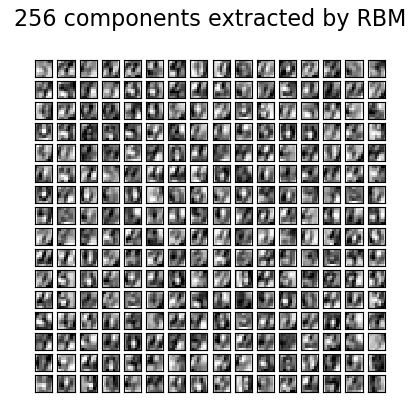
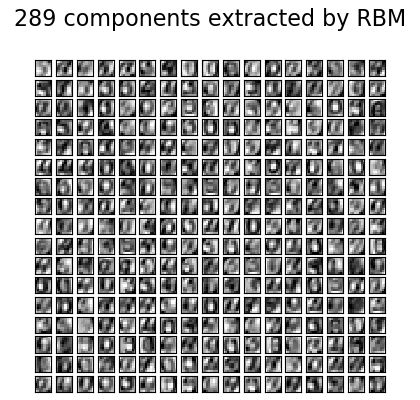
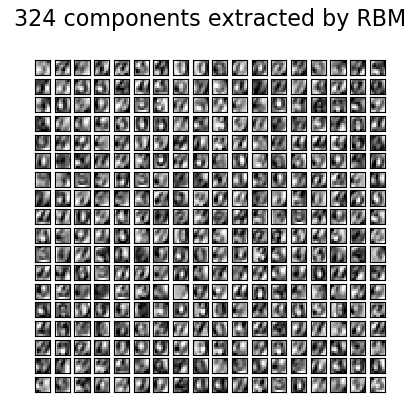
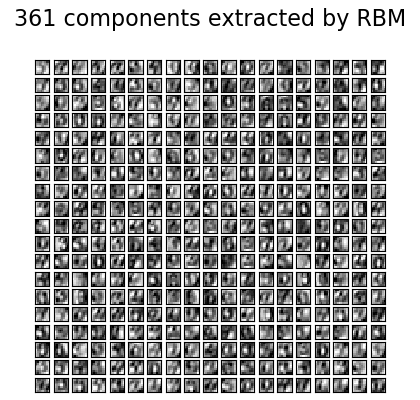
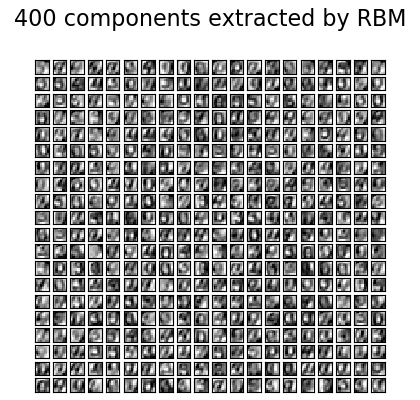
דביר גנץ 300852431

דרינה פויארקוב 316775402

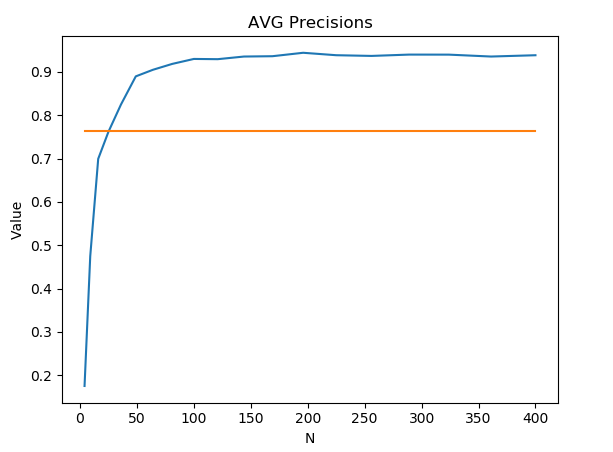
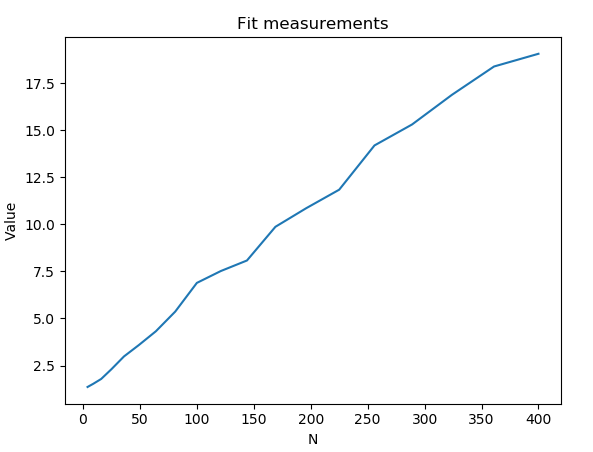
תרגיל 3

1.ב.

1. ג.

1.ד.1.

X\_train shape = (7188, 64)

X\_test shape = (1797, 64)

transform.X\_train shape = (7188, n\_component)

rbm.intercept\_hidden\_ shape = (n\_component, 1)

1.ד.2. הפונקציה mean\_hiddens\_ מתוך rbm.py:

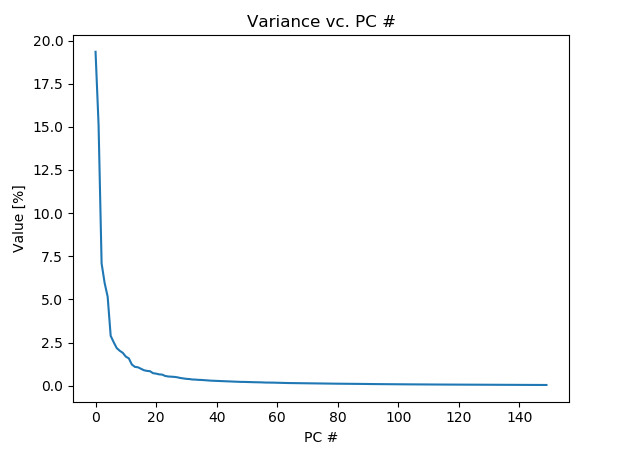
def \_mean\_hiddens(self, v):

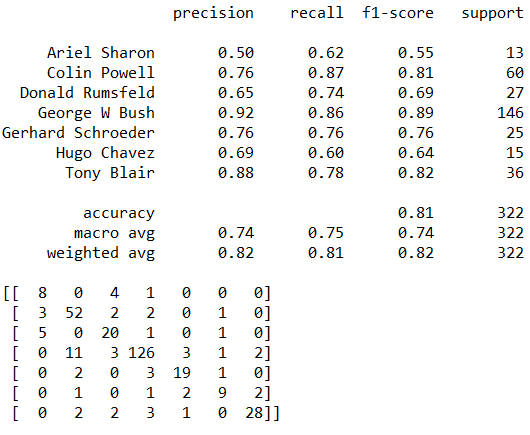
p = safe\_sparse\_dot(v, self.components\_.T)

p += self.intercept\_hidden\_

return expit(p, out=p)

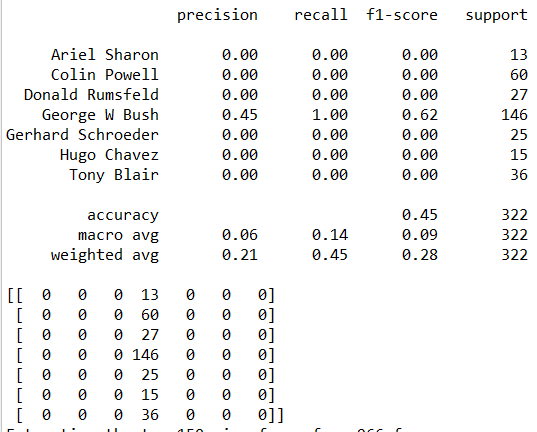
כאשר:  
self = RBM  
v = קלט, במקרה שלנו x\_train  
  
במונחים של המשתנים מהסעיף הקודם זה בעצם שקול ל:

2.א.   


ב. done in 17.066s  
  
  
  
  
ג-ד.





ה.   
done in 314.015s  


ו. ע"פ הנתונים שקיבלנו ניתן לראות בבירור מס' נקודות:  
- סיווג על נתונים במימד מופחת (S17~) מהירים משמעותית מסיווג על המימד המקורי (S314~).  
- רמת הדיוק של המסווג על המימד המופחת (0.74) גבוהה משמעותית מהדיוק על המימד המקורי (0.06).  
\* סיווג על המימד המקורי (1850) סיווג את כל התמונות כג'ורג' בוש.  
 (יתכן וזה קשור לעובדה שרוב התמונות בסט המבחן היו שלו)  
  
לסיכום התרגיל נסיק:   
- ככל שמימד הנתונים נמוך יותר זמן הריצה יהיה נמוך יותר  
- הפחתת מימדים יכולה לעזור לדיוק המסווג (כמו בשאלה 2) אבל לא בהכרח (כמו בשאלה 1.ג).  
יתכן וזה קשור למימד הנתונים המקורי אם הוא גבוה מאוד (1850 כמו בשאלה 2) חל עליו "קללת המימד הגבוהה" ולכן כדאי להפחית מימדים. לעומת זאת, אם המימד המקורי הוא לא מאוד גבוה (64 כמו בשאלה 1) אז לא כל כך כדאי להפחית את המימדים.