

Nama : Hafidh Fikri Rasyid

NIM : 1301142190

Kelas : IF-38-10

Assignment 2

Machine Learning

Cara Menjalankan Aplikasi :

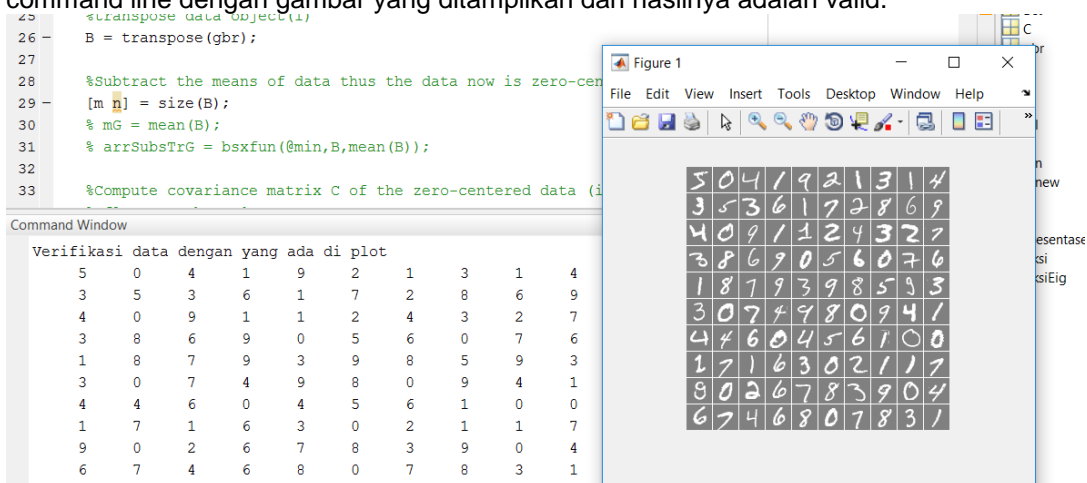
- Buka file PCA.m lalu klik run.

Penjelasan Aplikasi :

- Pada file PCA.m berisi baris kode pengerjaan soal a hingga d.

Jawaban :

- a. Pada bagian ini saya me-load data sebanyak 1000 data tetapi hanya ditampilkan sebanyak 100 data saja dan dilakuka verifikasi dengan mencocokkan data yang ditampilkan pada command line dengan gambar yang ditampilkan dan hasilnya adalah valid.



- b. Pada bagian ini ada beberapa langkah yang saya lewati tetapi salah satunya adalah nomor i dan ii karena saya menggunakan fungsi cov() pada matlab. Meskipun saya menggunakan fungsi cov(), saya juga mencoba mengerjakan secara manual tetapi untuk nomor iii dengan cara manual saya terdapat kesalahan perhitungan untuk nomor i,ii,dan iii dengan kode seperti dibawah ini :

Nomor i,ii dan iii(cara manual) :

```
%transpose data object(i)
B = transpose(gbr);

%Subtract the means of data thus the data now is zero-centered.(ii)
[m n] = size(B);
mG = mean(B);
arrSubsTrG = bsxfun(@min,B,mean(B));

%Compute covariance matrix C of the zero-centered data (iii)
f1 = arrSubsTrG';
C = (arrSubsTrG * f1)/(n-1);
```

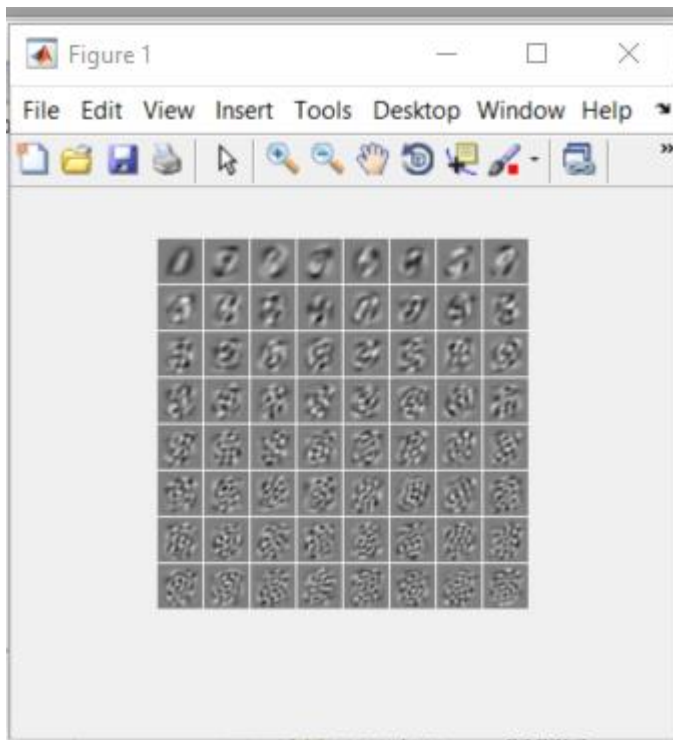
Nomor Nomor i,ii dan iii(cara menggunakan fungsi) :

```
C = cov(gbr);
```

Untuk pencarian vector eigen saya menggunakan fungsi eig() dimana dari fungsi tersebut menghasilkan dua data array yaitu nilai eigen dan vector eigen. Cara penggunaanya dalah adalah seperti kode dibawah ini (soal iv):

```
[v L] = eig(C);
```

Setelah itu seleksi descending eigen values berdasarkan hasil sort dari vector eigen sebanyak 64 buah. Untuk sorting saya menggunakan metode bubblesort. Setelah itu visualisasi data sebanyak 64 buah tersebut dan menghasilkan sebuah visual gambar seperti dibawah ini (soal v dan vi):



Untuk menghasilkan domain PCA yaitu dengan mengaplikasikan rumus dibawah ini :

$$Y = v_{row} * B$$

Dimana vrow adalah data 64 nilai eigen terbaik yang ditransposkan dan B adalah nilai dari gambar dataset yang berbentuk array dan diimplementasikan dengan kode dibawah ini :

```
Y = slksiEig'*B;
```

Dari hasil perhitungan diatas maka didapat sebuah matrix yang berukuran 64x1000

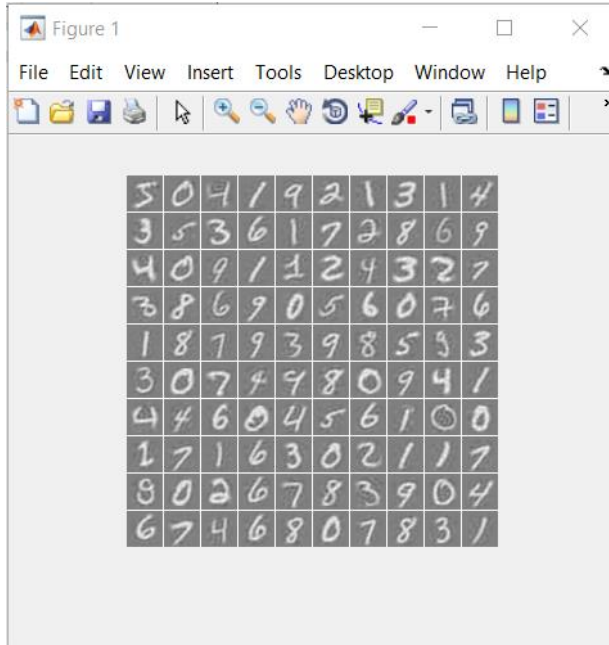
- c. untuk mereconstruct gambar menggunakan dari domain PCA maka digunakan rumus

$$B_{rev} = v_{row}^T * Y$$

Dimana vrow adalah data 64 nilai eigen terbaik dan Y adalah domain PCA. Untuk implementasi kodenya sebagai berikut:

```
Bc = slksiEig * Y;
Bct = Bc';
visual(Bct(1:100,:));
```

Dan hasil dari kode diatas adalah :



Dari sini dapat disimpulkan bahwa data dari domain PCA dapat direconstruct atau dikembalikan ke bentuk semula dengan mengalikannya dengan data terbaik yang telah dipilih. Meskipun begitu data telah direconstruct tidak identik dengan data sebenarnya ini dibuktikan dengan tingkat akurasi yang dihasilkan pada jawaban soal dibawah.

- d. Untuk menghitung persentase data maka digunakan rumus jumlah dari nilai vector eigen terbaik sebanyak 64 buah dibagi jumlah dari seluruh nilai vector eigen dengan mengaplikasikan kode dibawah ini :

```
a1 = sum(slksi);
a2 = sum(aSort);

presentase = (a1(2)/a2(2))*100;
```

Dan dari situ bisa kita hitung persentase correct reconstruction yaitu sebesar 87,9252%

Presentase dari correct reconstruction adalah 87.9252%