Ali helmut

1301154246

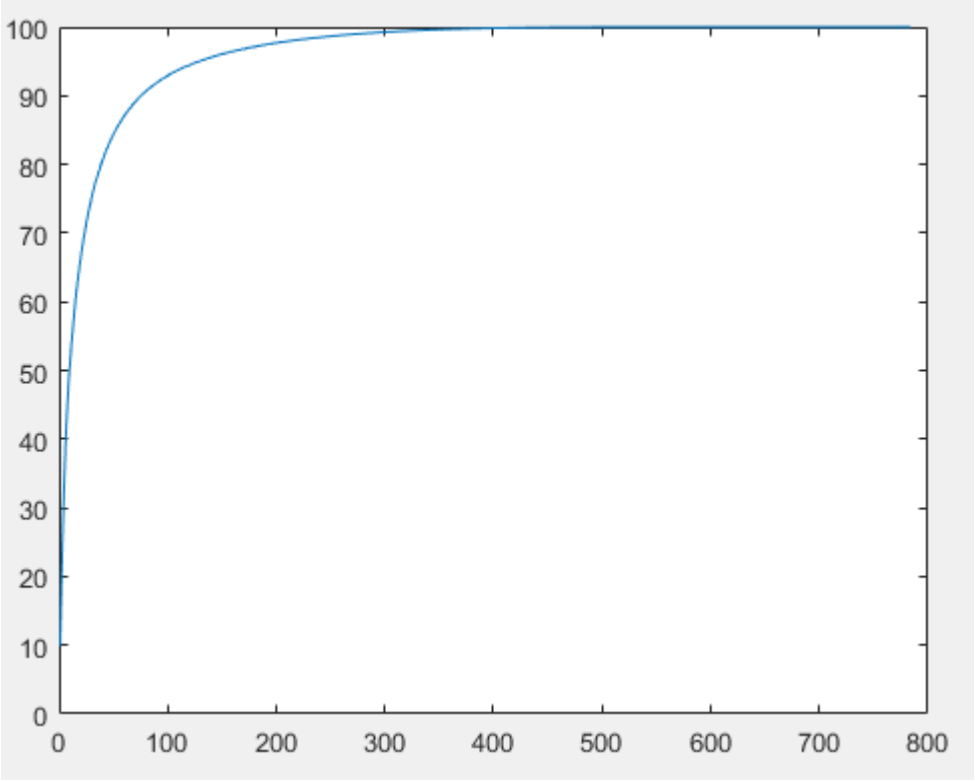
- E. Compute the percentage of correct reconstruction for each of k first principal components, k ∈ [1..784]. Plot the result in a figure where x-axis corresponds to variances λi and y-axis corresponds to percentages of correct reconstruction

```
nilaiHasilRekontruksi2 = [];
for i = 1 : 784
    a = 0;
    for j = 1 : i
        a = a + nilaiCDesc(j);
    end
    c = (a/b) * 100;
    nilaiHasilRekontruksi2 = [nilaiHasilRekontruksi2; c];
end
plot(nilaiHasilRekontruksi2);
```

Keterangan : B adalah jumlah dari nilai eigen (didapat dari proses sebelumnya)

```
for i=1 : 784
    b = b + nilaiCDesc(i);
end
```

Dan hasil plot sebagai berikut :



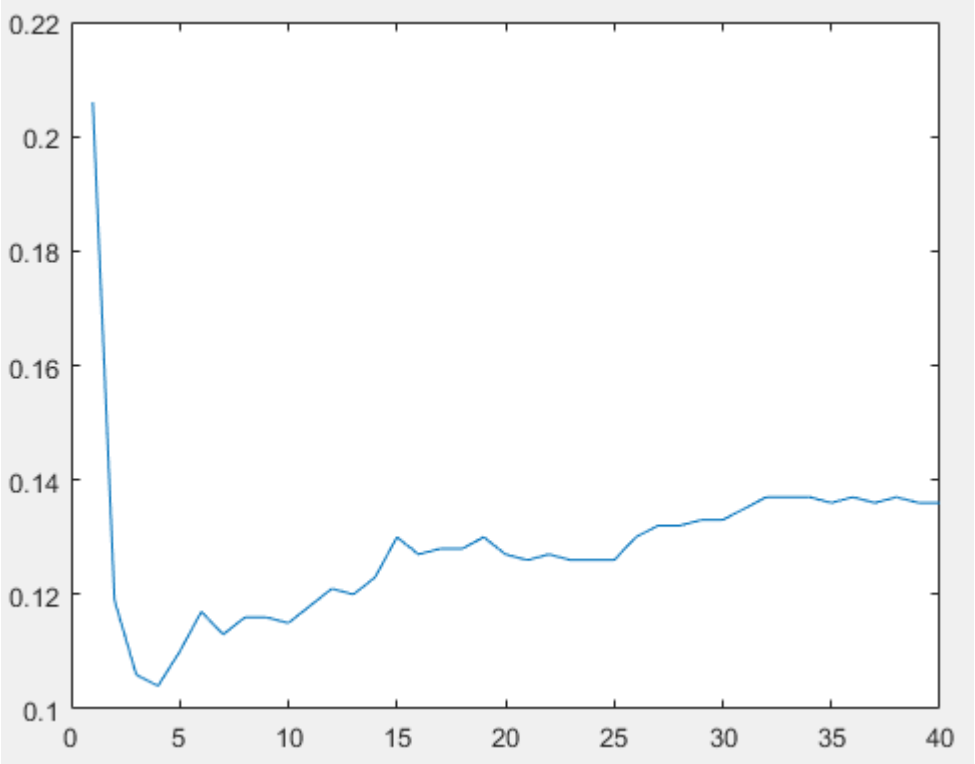
- F. Define m ∈ [5..200] with interval 5. For each m first principal components do following steps, [12 points]
  i. Project the loaded 1000 data into eigen subspaces (PCA domain).
  ii. Classify data on PCA domain using kNN with k = 15 (hint: it is similar with classifying on training set.)
  iii. Calculate the misclassification rate. (Note: for simplification, misclassification rate is the number of wrong prediction divided by total number of data.)

```
missRate = [];
 for i=1 : 40
    Vrow                = VectCDesc(1:i*5,:);
    PCADomain           = Vrow * B;
    PCADomain           = PCADomain';
    hasilClasifikasi    = [];
    a = 0;
      for j = 1 : 1000

        jarak = [];
        for k = 1 : 1000
            jarak = [jarak; norm(PCADomain(j,:) - PCADomain(k,:)),y(k)];
        end

        jarak             = sortrows(jarak,1);
        hasilClasifikasi  = [ hasilClasifikasi; mode(jarak(1:15,2)) ];
        a                 = a + (hasilClasifikasi(j) ~= y(j));
      end

      missRate = [ missRate;  a/ 1000 ];
 end
```

missRate 40x1 double

- G. From f(iii), we have 40 numbers of misclassification rates obtained from each m first principal components. Using line plot, plot the all misclassification rates into a figure where x-axis corresponds to m first principal components and y-axis corresponds to misclassification rates

```
plot(missRate);
```



- H. According the line plot on g, what is the m first principal components that gives the best rate?

  Berdasarkan hasil plot di atas dapat dilihat error rate terendah berada pada rentang sekitar 3-4 (dikalikan 5, maka sekitar 15-20 first principal components)