In The name of God

University Tehran

Engineering Facility

Electrical and Computer Engineering

Pattern Recognition

HW # 2.2

Amin Fadaeinedjad

810195442

Spring 98

List

Abstract

This homework is about using the Estimation maximization and the hidden Markov Model. At the first question of this homework we have a sample of data from a number of Normal distribution with different mean and variance and we add them with each other and we have a mixture of normal distribution we mission is to without knowing the mean and variance of the distribution we should estimate the mean and variance of them by using EM this works by improving the mean and the variance in every step and after about 40 steps of the feedback we can see the result.

The next two part we are using the HMM to give the best probability for a certain output. It work like we only can see the observation of a state and we only have the percentage of a change.

Question 1

In this part we have 5 normal distribution with a certain mean and variance and we have generated 1000 data from 5 different norms and we added them to each other the mission is to calculate the mean and variance of the mixture. We are using the EM algorithm to estimate the norm parameters.

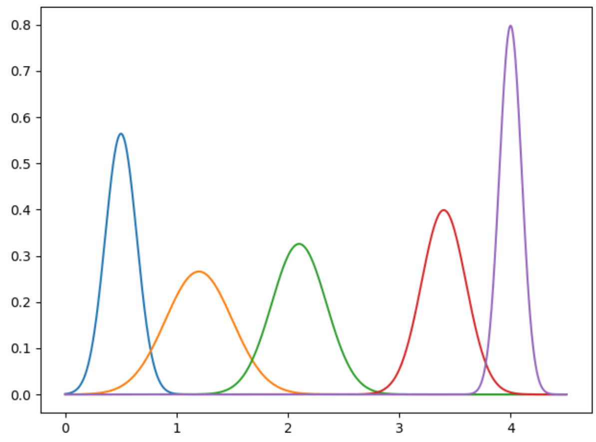
$$P(x_i|class_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{\frac{1}{2}(x_i-\mu_1)^2}{\sigma_1^2}}$$

$$prob(x_i|class_1) = \frac{P(x_i|class_1)}{P(x_i|class_1) + P(x_i|class_2) + P(x_i|class_3) + P(x_i|class_4) + P(x_i|class_5)}$$

$$\mu_j = \frac{\Sigma\left(prob(x_i|class_j) * x_i\right)}{\Sigma\left(prob(x_i|class_j)\right)}$$

$$\sigma_j^2 = \frac{\left(\Sigma(x_i - \mu_j)^2 * prob(x_i|class_j)\right)}{\Sigma\left(prob(x_i|class_j)\right)}$$

And by repeating this for several times we start with a prior distribution with a particular mean and variance.



The real distribution of the norms

Figure 1

a) In this part we are using 5 class to estimate the distribution the data set is like Figure1

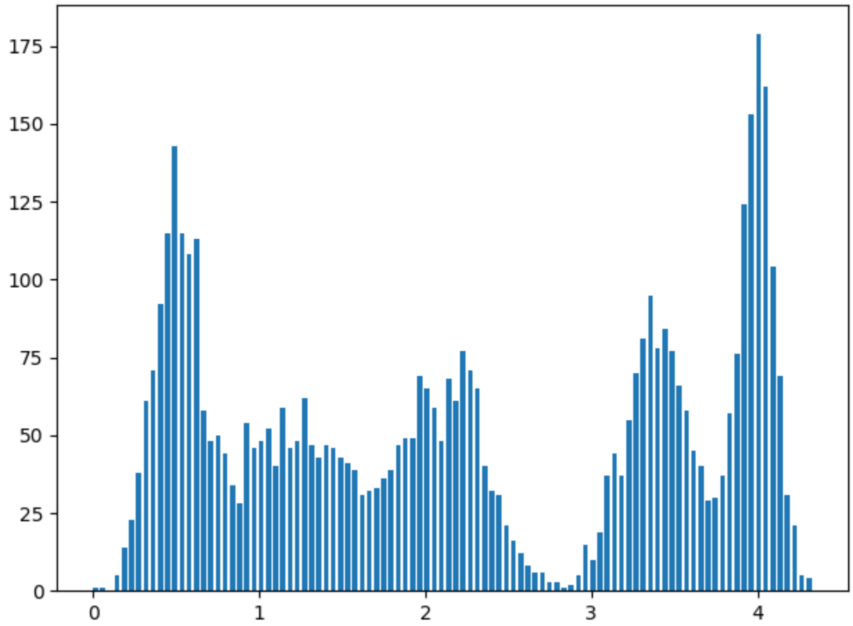
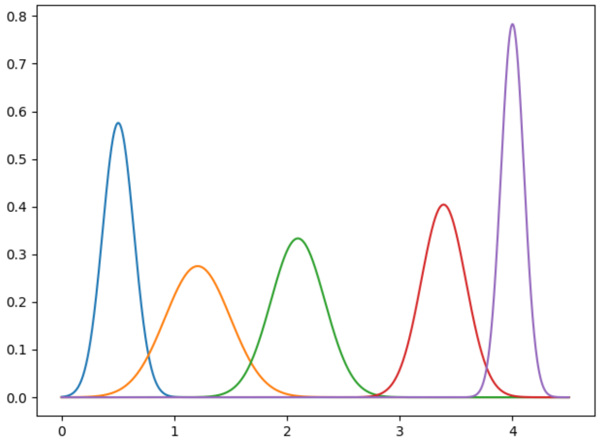


Figure 2 Histogram

According to Figure2 we show the data from the 5 distribution and 5 class for classifying the data the output is something like this.



After 40 times training the machine

Figure 3 EM estimate

We started with a prior mean and variance $\mu = [0.5, 1.1, 2.3, 3.1, 4.0]$ $\sigma^2 = [0.5, 0.05, 0.5\ 0.5\ 0.5]$ and after calculating 50 times with the mentioned algorithm and got figure 3 as the output.

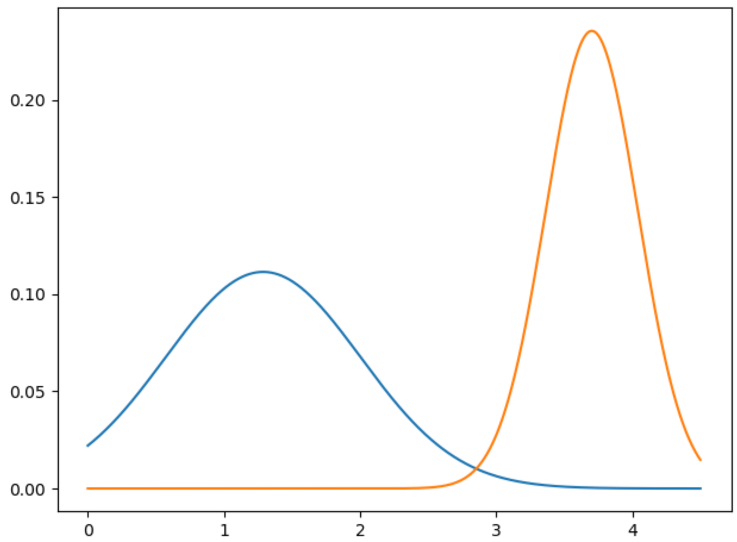
The new mean for this is:μ:

$[0.4914995648427, 1.19211185183, 2.0970024753, 3.3871390147, 4.0010829227]$

The new variance is:σ:

[0.0194713768729, 0.080607602, 0.061007588403, 0.03609639761, 0.0099469136291

b) In this part we are going to use only two class of classify the data set given by the 5 Gaussians the result will something similar to Figure 4



The result for 2 class

Figure 4

The prior mean and the variance given in this part is $\mu = [1.8\ 3.5]$ $\sigma = [1\ 1]$

And the final result that we have received from the machine was

The mean of the result: μ:

[1.2880016317400824, 3.7027499538280235]

The variance of the output is: σ:

[0.5121726927703674, 0.11477087304937253]

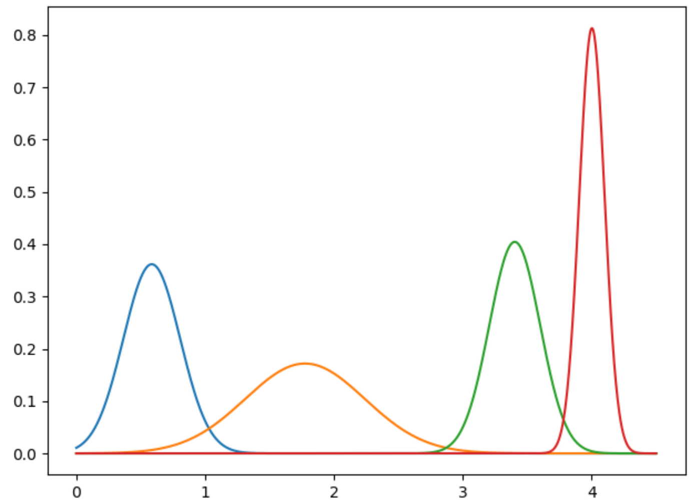
We can compare this with the previous part because we have less class for estimating the output we are going to have more errors compared to 5 classifier so is this case the norms are going to fit the previous norms.

c) In this part we are going to reduce the number of norms that are going to fit the data we have in the past we used 5 to cover the data in this part we used 4 the result was Figure 5.
The prior mean and variance that we have is:

$$\mu = [1.2\ 2.5\ 3.4\ 6]$$

The prior variance that we have for the distribution is:

$$\sigma = [1\ 1\ 1\ 1]$$



There are 4 norms in it

Figure 5

In this new design that there are only 4 Gaussian distributions with a certain parameters.

The new mean for this machine is:

$[0.5860868530199983, 1.7804583322477645, 3.4035085335188997, 4.001671854149498]$

And the new variance for the new distribution is:

[0.055626329, 0.22467754303, 0.034328066826, 0.009936881796]

This part we have 4 class to classify the data that we have but there are data's that we can't fit them properly we one of the distributions is between two other distributions this has more error than the 5 class classifier but has less error than the two class classifier.

d) In this part we are going to see how accurate does the machine work with a less amount of data by seeing the image of the new distributions we can find out that if we have more data of the distributions the better predictions we are going to have.

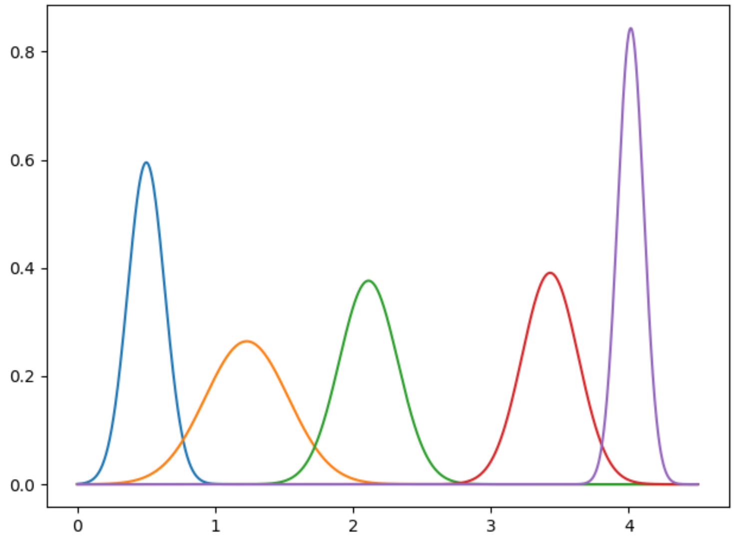


Figure 6

The mean of the output is:

$$\mu = [0.4996561459, 1.228635060, 2.112264842, 3.430542677, 4.01562160050]$$

And the variance of the estimation is:

$$\sigma = [0.01798211823, 0.09109658220, 0.0449522161, 0.041663676, 0.008952939]$$

When we compare this with the other classifiers we can see that when we increase the number of data that we have a better classifier compared to the classifier with less data.

Question 2

This part we are going to use Markov hidden model to solve the best answer for the case that we have observed from it.

We have two case PESESEPEPSP and SESPSEPSPSE and we are going to use forward selection algorithm to give the probability for this observation.

We have a beginning state that is the probability that each one can happen and by reading the question the probability is $\pi = [0.6, 0.4]$ which is the beginning chance and we have the chances of the observations that can occur for sleep= [0.1, 0.6] poop= [0.4, 0.3] eating= [0.5, 0.1] which the first index in showing the probability that it's healthy and the second index is the probability of being sick.

And at the last part we have a matrix that calculates the chances after changing state from healthy to sick and other cases of changes and all of this can be modeled to a matrix the following matrix is called the change = $\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$ and we are going to us this algorithm to calculate the chance of the healthy or the sickness of the dog.

The result is lick this if the observations was EPPS:

$$result = \pi .* eat * change .* poop * change .* poop * change .* sleep$$

Which the result vector is a 1*2 dimensional vector that the first index is the chance of being healthy and the second index is the chance of being sick.

We have done this is code by python and they are attached to the following homework.

For the input: PESESEPEPSP the probability is:

[1.88501712e-06 1.63492826e-06]

Which means that the chance of being healthy is a bit more than the chance of being sick so in the first case we can admit that the dog is probably healthy

For the input: SESPSEPSPSE the probability is:

[1.98206902e-06 4.79557815e-07]

In this case we can see that just like the previous part the chance of getting healthy in more than the chance of being sick in it so the answer is that the dog is sick.

And the end we learned that we can estimate the output by computing the probability of the output other which it is healthy or sick depending on the observations that we have by the pattern of the observations.

Question 3

1)

$$P(q_t|q_{t+1},\dots,q_T) = \frac{P(q_t,q_{t+1},\dots,q_T)}{P(q_{t+1},q_{t+2},\dots,q_T)} \to \frac{P(q_T|q_t,q_{t+1},\dots,q_{T-1})P(q_t,\dots,q_{T-1})}{P(q_{t+1},\dots,q_T)}$$

Be knowing the Markov algorithm = $\frac{P(q_T|q_{T-1})P(q_t,\dots,q_{T-1})}{P(q_{t+1},\dots,q_T)}$ =and keep on using the Markov probability we will reach a point that.

$$= \frac{\prod_{i=0}^{T-t-1} P(q_{T-i}|q_{T-i-1})P(q_t)}{\prod_{i=0}^{T-t-2} P(q_{T-i}|q_{T-i-1})P(q_{t+1})} = P(q_{t+1}|q_t) * \left(\frac{P(q_t)}{P(q_{t+1})}\right) =$$

$$\frac{P(q_{t+1},q_t)}{P(q_{t+1})} = P(q_t|q_{t+1}) \to$$

$$P(q_t|q_{t+1},\dots,q_T) = P(q_t|q_{t+1})$$

2) Model 1 in this model just like the previous part we have a beginning vector called the π vector after $\pi = [0.4, 0.6]$ $A = [0.45, 0.5]$ $B = [0.55, 0.5]$

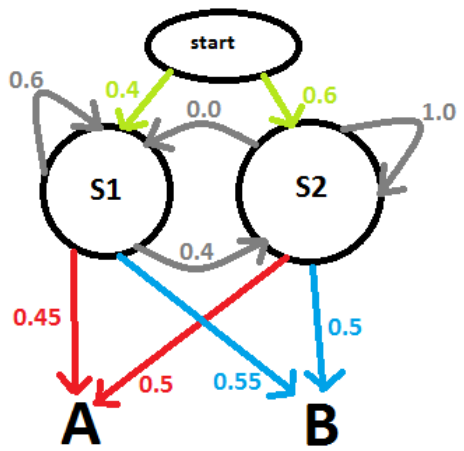
$$change = \begin{bmatrix} 0.6 & 0.4 \\ 0.0 & 1 \end{bmatrix}$$

For Model 2 is like this

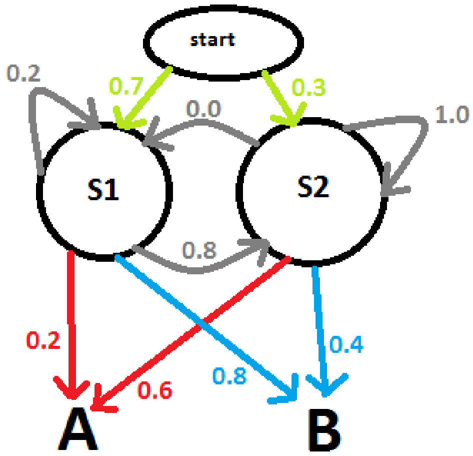
$$\pi = [0.7, 0.3] \quad A = [0.2, 0.6] \quad B = [0.8, 0.4]$$

$$change = \begin{bmatrix} 0.2 & 0.8 \\ 0.0 & 1 \end{bmatrix}$$

The Model 1 state diagram:



The state diagram of Model 2 is like this:



b) In this part we are going to calculate the chance that the sequence {A,B,A} happens we are going to use the same algorithm from the question 2 which we calculate what is the chance to see the sequence and it is equal to:

$$P(Sequence) = P(Sequence|S1) + P(Sequence|S2)$$

And we have the two amounts that we need to calculate we used python code to generate the chance that we reach the last sequence seeing S1 or S2 and by just summing them we get $P(Sequence)$

For model 1 after $\pi.* A * change.* B * change.* A$ we have this as the output:

$$[0.016038, 0.10488\,] \rightarrow sum \rightarrow 0.120918$$

For model 2 after $\pi.* A * change.* B * change.* A$ we have this as the output:

$$[0.000896\ 0.080832] \rightarrow sum \rightarrow\ 0.081728$$

So we can see that the chance of getting the ABA sequence in model 1 is more than the chance of getting the sequence in model 2

So model 1 is more likely to see that sequence to model 2

Process

In this homework we coded in python to see the maximum chance for the output to have a certain observation and the codes are in the zipped file to see.

And there are some codes for part 1 on this HW that are based on the EM estimarion.

Reference

Automatic speech recognition a deep learning approach

stackoverflow.com

youtube.com

https://scholar.google.com/scholar?q=Forward+Algorithm+using+hidden+markov+model&hl=fa&as_sdt=0&as_vis=1&oi=scholart

https://eu.udacity.com/course/introduction-to-computer-vision--ud810