Homework set #2 - Optimization, GMM

Ido B. Gattegno, Haim H. Permuter

August 31, 2016

1 Optimization

- Read Chapter 1 in Boyd book on convex optimization [1].
  Free online version: `https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf`.
- Read Chapter 2.1-2.3 and solve 2.1-9
- Read Chapter 3.1-3.2 and solve 3.1-3.10
- Read Chapter 4.1-4.3 and solve 4.1-4.10
- Read Chapter 5.1-4.2 and solve 5.1-4.5

2 Gaussian Mixture Models

Gaussian Mixture Models (GMMs) are statistic models, i.e., hypothesis on the behavior of the data. The family of GMMs are defined by various of parameters, such as the number of Gaussians in a mixture, means, covariances, etc. Each set of parameters defines a statistical model, which helps us classify samples, by giving the probability of a sample given the parameters.

In machine learning we use Expectation Maximization (EM) to train some of the GMM parameters for each class. However, not all of the parameters are learned. Some parameters are given in advance, such as the number of Gaussians in each model. There are techniques to learn what is the best parameter. However, it is strongly recommended to try a small set of parameters before performing optimization.

In the following implementation tasks, take time at the beginning of the learning algorithm until the end. Do not include data generation, feature extraction/reduction or decision making. During implementation, it is important to check that the algorithm runs in a reasonable time.

2.1 Learning synthetic data

This section will help you build and learn the behavior of GMMs. It is highly recommended to make your implementation readable and compatible for generic uses. It is likely that you will use these implementation in the future.

1. Define Gaussian mixture models in order to generate data. Use at least 2 classes, in each class use 3 clusters. The data should be generated randomly, both the latent variable and the samples. The data dimension should be at least 2.
2. Implement the K-Means algorithm and test it on the generated data.
   You may use scatter plots to visualize the data and the result from the clustering.
3. Try various ways to examine the behaviour of K-Means.
   For instance, try various options for K, change the means of the data, etc.

4. Implement expectation maximization for GMMs. Your implementation should support both diagonal and full covariances.

5. Test your models on the synthetic data that you learned.
   For your tests, use K-Means to train the initial model for each class.
   Note: K-Means results in clusters centroids and samples assignments. However, for initial models we need weights, means and covariances. Use the results from K-Means in order to estimate these parameters.

6. There are more techniques to initialize EM. You may read about Universal Background Models (UBM) in [2],[3], or other techniques. GMMs are widely used for speaker verification.

2.2 MNIST database

Note: Make a short report on the following experiments.

We will first try our models on raw data of MNIST database.
Gaussian mixture models use empirical covariance matrix for estimation.
When working with raw data, you may experience problems with the algorithm, due to singularity of covariance matrix.
A way to work around the problem, is to perform *diagonal loading*. First, check the condition number of the covariance matrix. Then, if it too large (for example larger than 10^3), add an identity matrix to the covariance. Due to that, the condition number will reduce, i.e., the eigen values will increase.
Another issue that may come out, is that the features (in this case pixels) are highly dependent. It may cause the EM algorithm to fail. If this is the case, you may want to try a method for dimension reduction.

- Use your implementation of the GMMs to learn and test digits from the MNIST database.
  Do not process the data, use raw images as observations. Try various parameters. What are your results?
- Try various methods for initialization.
- Process the data in order to reduce dimensionality, using Principal Component Analysis (PCA). Be sure that the PCA coefficients are determined only by training set, and not the test.
- Test GMMs on the new data. Use the methods you learned, and try various parameters.
- Measure your success and compare the methods you implement. Be sure to include error rate and the good parameters for each method you tested.
- Learn about feature extraction methods for this database. Do not spend a lot of time on those methods, we aim to focus on learning and not image processing. You may read about it in [4].
- Compare the resuts to the K nearest neighbour.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1):19 – 41, 2000.

[3] Haim Permuter, Joseph Francos, and Ian Jermyn. A study of gaussian mixture models of color and texture features for image classification and segmentation. *Pattern Recognition*, 39(4):695–706, 2006.

[4] Rafael MO Cruz, George DC Cavalcanti, and Tsang Ing Ren. Handwritten digit recognition using multiple feature extraction techniques and classifier ensemble. In *17th International Conference on Systems, Signals and Image Processing*, pages 215–218, 2010.