E9 205 – Machine Learning for Signal Processing

Homework # 1
Due date: Aug 31, 2016 (3:30PM)

1. **Sparse Non-negative Matrix Factorization**.
The non-negative matrix factorization (NMF) is the problem of factorizing a non-negative matrix $\mathbf{V}$ of dimension $n \times m$ to non-negative factors $\mathbf{B}$ of dimension $n \times r$ and $\mathbf{W}$ of dimension $r \times m$. In an over-complete basis formulation, r is chosen to be high and the weights are intended to be sparse. The sparsity of the weights is imposed by minimizing the L_1 norm of columns of $\mathbf{W}$ defined as $||\mathbf{W}||_1 = \sum_{jk} |w_{jk}|$ where w_{jk} is the j,k th element of $\mathbf{W}$ with $j \in \{1, .., r\}, k \in \{1, .., m\}$. In this case, the loss function to be minimized becomes

$$L = Div(\mathbf{V}, \mathbf{BW}) + \lambda ||\mathbf{W}||_1$$

   (a) Find the suitable auxiliary function and show that it satisfies the properties of auxiliary functions.

   (b) Show that the update equations for k th column of $\mathbf{W}$ containing elements w_{jk} with $j \in \{1, .., r\}$ are,

$$w_{jk}^{t+1} = \frac{w_{jk}^{t}}{\sum_{i=1}^{n} b_{ij} + \lambda} \sum_{i=1}^{n} \frac{v_{ik}}{\sum_{p=1}^{r} b_{ip} w_{pk}^{t}} b_{ij}$$

   where b_{ij} is the i,j th element of $\mathbf{B}$ with $i \in \{1, .., n\}, j \in \{1, .., r\}$.

   (c) Find the update equation for i th row of $\mathbf{B}$ containing elements b_{ij} with $j \in \{1, .., r\}$ (Hint - $\mathbf{V}^T = \mathbf{W}^T \mathbf{B}^T$).

   Remark- Analytical ! (Points 5 + 10 + 10)

2. **Properties of Linear Prediction**.

   (a) Let $\mathbf{R}_N$ denote the Hermitian Toeplitz autocorrelation matrix of size N. Using the augmented normal equations for order N predictor, show that the energy of the prediction error signal $\mathcal{E}_N$ is given by

$$\mathcal{E}_N = \frac{det(\mathbf{R}_{N+1})}{det(\mathbf{R}_N)}$$

   where *det* denotes the determinant of the matrix.

(b) Consider the linear prediction of a real WSS process $x[n]$ with autocorrelation $R(k)$. Without using the orthogonality property of prediction error, show that the energy of prediction error $\mathcal{E}_N$ can be written as

$$\mathcal{E}_N = R(0) + \mathbf{a}^T \mathbf{R}_N \mathbf{a} + 2\mathbf{a}^T \mathbf{r}$$

where $\mathbf{r} = [R(1)\ R(2)\ ...\ R(N)]^T$ and $\mathbf{a} = [a_{N,1}\ a_{N,2}\ ...\ a_{N,N}]^T$. Show that by differentiating $\mathcal{E}_N$ w.r.t $\mathbf{a}$ and setting it to $\mathbf{0}$, we can obtain the normal equations. (Hint - differentiating a scalar with a vector - $\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$ and $\frac{\partial \mathbf{x}^T \mathbf{y}}{\partial \mathbf{x}} = \mathbf{y}$).

Remark- Straight forward ! (Points 5 + 10)

3. **Implementing NMF** - Implement the NMF algorithm (divergence cost) to estimate the basis $\mathbf{B}$ and weights $\mathbf{W}$ given a positive matrix $\mathbf{V}$. Apply the NMF for

   (a) Face images - 10 subjects with normal lighting (2 images per subject) are here. *http://leap.ee.iisc.ac.in/sriram/teaching/MLSP/assignments/HW1/images.zip* Each gif image is 100×100 matrix. This can be converted to a vector of dimension 10000×1. Using the NMF on training data ($\mathbf{V}$ is of dimension 10000×20), estimate the basis and weight matrices. Show that the divergence is a non-increasing function as the number of iterations is increased. For each test subject, will the basis vectors suffice to reconstruct the partial occulsion found in right/left lighting conditions provided in the test images. How many basis vectors would be a good choice ? Are the observations made valid for both test subjects ? Illustrate using plots.

   (b) Speech spectrogram - We have clean, noise and noisy speech files here *http://leap.ee.iisc.ac.in/sriram/teaching/MLSP/assignments/HW1/speech.zip* The files are in wav format sampled at $16kHz$. Using the power spectrogram of speech files (use 25 ms Hamming windows with a shift of 10 ms for spectrogram computation), compute NMF basis for clean and noise file. For a speech file with 100 frames and 256 point power spectrum, $\mathbf{V}$ is of dimension 256×100. Show that the divergence is a non-increasing function as the number of iterations is increased. The same NMF can be applied on the noise file as well. With the basis functions from clean and noise files, is it possible to separate speech and noise from the noisy speech file spectrogram. How many basis functions are needed. Illustrate the denoising process using plots.

   Submit the code (in your choice of language) along with the report.

   Remark- First coding fun ! (Points 20 + 20)

4. **Implementing LP Spectrogram** - Implement the LP algorithm on each short-term frame (25 ms) of speech file (autocorrelation method). Show that as order of LP is increased, the energy of the error signal is non-increasing. Using the clean and noisy speech files provided above, in each short-term window, compute the LP coefficients and the LP power spectrum. Plot the LP spectrogram for clean and noisy speech file. Comparing the conventional power spectrogram with LP spectrogram, which one of them is visually more appealing in terms of having similar representation from clean and noisy file. How does the order of LP affect these plots.

   Remark- Feeling good at coding (Points 20)