# E9 205 – Machine Learning For Signal Processing

Practice Exam

Date: Dec. 1, 2017

**Instructions**

1. This exam is open book. However, computers, mobile phones and other handheld devices are not allowed.
2. Notation - bold symbols are vectors, capital bold symbols are matrices and regular symbols are scalars.
3. Answer all questions.
4. Total Duration - **180 minutes**
5. Total Marks - **100 points**

Name - ................................

Dept. - ....................

SR Number - ....................

1. **Text analysis** - In document analyis, the number of co-occurences $n(d_i, w_j)$ of word $w_j$ in a document $d_i$ is obtained for all words $j = 1, .., M$, and all the documents $i = 1, .., N$, where $M$ is total number of words in the vocabulary and $N$ is total number of documents. The assumption in this analysis is that there is an underlying topic $z_k$ for $k = 1, .., K$ for each of the document. The joint probability model in this case is,

$$P(d_i, w_j) = P(d_i)P(w_j|d_i) = P(d_i)\sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i)$$

This forms a generative model where a document is selected with probability $P(d_i)$, a topic is then selected with probability $P(z_k|d_i)$ and word is generated with a probability $P(w_j|z_k)$. The total log likelihood of the co-occurence model is given by,

$$\begin{aligned}\mathcal{L} &= \sum_{i=1}^{N}\sum_{j=1}^{M} n(d_i, w_j)\log(P(d_i, w_j)) \\ &= \sum_{i=1}^{N} n(d_i)\left[P(d_i) + \sum_{j=1}^{M}\frac{n(d_i, w_j)}{n(d_i)}\log\left\{\sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i)\right\}\right]\end{aligned}$$

where $n(d_i) = \sum_{j=1}^{M} n(d_i, w_j)$ is the document length. Formulate and solve for the unknown probability mass functions $P(w_j|z_k)$ and $P(z_k|d_i)$ using the EM algorithm. (**Points 20**)

2. **Convolutional Networks** - A CNN realizes a convolution operation of input image $\mathbf{X}$ of size $(U, V)$ with a set of weights (filters) $\mathbf{W}^k$ for $k = 1, .., K$ where $K$ denotes the number of filters in a CNN layer. The convolution operation is given by,

$$\begin{aligned}\mathbf{Y}^k &= \mathbf{X} * \mathbf{W}^k \\ \mathbf{Y}^k(p, q) &= \sum_{i=0}^{S-1} \sum_{j=0}^{T-1} \mathbf{X}(p+i, q+j)\mathbf{W}^k(i, j)\end{aligned}$$

where $(S, T)$ is the size of the filter $\mathbf{W}^k$, $p$ ranges from $0, 1, ..., U - S$ and $q$ ranges from $0, 1, ..., V - T$. Note that the output image $\mathbf{Y}^k$ is of size $(U - S + 1, V - T + 1)$. Let $J$ denote the cost function used in CNN training. Assume that the partial derivative w.r.t. to output of filter has been computed as $\frac{\partial J}{\partial \mathbf{Y}^k}$. Prove the following gradient update rule for filter learning

$$\frac{\partial J}{\partial \mathbf{W}^k} = \mathbf{X} * \frac{\partial J}{\partial \mathbf{Y}^k}$$

(**Points 20**)

3. **Restricted Boltzmann Machine** - A modified Gaussian RBM is defined using visible units $\mathbf{v}$, hidden units $\mathbf{h}$ with the energy function and the joint probability density function given by,

$$
\begin{aligned}
E(\mathbf{v}, \mathbf{h}) &= 0.5 \sum_{i \in visible} \frac{|(v_i - b_i)|^2}{\sigma_i^2} - 0.5 \sum_{j \in hidden} \frac{|(h_j - c_j)|^2}{\sigma_j^2} - \sum_i \sum_j \frac{v_i h_j}{\sigma_i \sigma_j} w_{ij} \\
P(\mathbf{v}, \mathbf{h}) &= \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{Z}
\end{aligned}
$$

where $Z$ is a normalization constant and $v_i, b_i$ are the $i$th dimension of visible layer $\mathbf{v}$ and bias vector $\mathbf{b}$ respectively and $h_j, c_j$ are the $j$th dimension of hidden layer $\mathbf{h}$ and bias vector $\mathbf{c}$ respectively. The parameters $\sigma_i, \sigma_j$ are scaling constants of the visible and hidden layer respectively. For this RBM definition,

(a) Show that conditional probability density function $p(v_i|\mathbf{h})$ is Gaussian. What are the parameters of the Gaussian distribution. (**Points 10**)

(b) Find the conditional probabilty density function of the hidden unit $h_j$ given the visible layer input $\mathbf{v}$, i.e., $p(h_j|\mathbf{v})$. **(Points 10)**

4. **Line Mixture Model** A line mixture model is the problem of fitting a mixture of lines on a 2-D dataset. Let $\mathbf{z}_i = [x_i \; y_i]^T$ denote a set of 2-D data $i = \{1, .., N\}$. Each mixture component in the LMM is defined using a line $f_k(x_i) = a_k x_i + b_k, \; k = \{1, ..., K\}$, where $K$ is the number of mixtures and $a_k, b_k$ are the parameters of the line for the $k$th mixture component. The pdf of $z_i$ is modeled as,

$$p(z_i|\lambda) = \sum_{k=1}^{K} \alpha_k \mathcal{N}(y_i; f_k(x_i), \sigma_k^2)$$

where $\sigma_k$ is the variance of the $k$-th mixture component and the model parameters $\lambda = \{a_k, b_k, \sigma_k\}_{k=1}^{K}$. Given a set of $N$ data points,

(a) Write down the Q function which will allow the EM estimation of the $\lambda$.

(b) Find the iterative maximization steps for all the parameters in the model.

**(Points 10)**

5. **MLSP Exam and grading** - Prof. Sam is evaluating the final exam of his MLSP course which was taken by $N$ students. The exam had $Q$ questions. From the answers provided by students, he finds the assignment variable $x_{nq}$ where $(x_{nq} = 1)$ indicates that the answer for student $n$ and question $q$ was correct and $(x_{nq} = 0)$ indicates answer for student $n$ and question $q$ was incorrect. Here $n \in \{1, .., N\}$ and $q \in \{1, .., Q\}$. Each question is assigned a latent difficulty $\delta_q$ and each student is associated with a latent ability $\alpha_n$. Prof. Sam uses a sigmoidal model for the conditional probability of the assignment variable $(x_{nq} = 1)$ given the latent ability vector $\boldsymbol{\alpha} = [\alpha_1, .., \alpha_N]^T$ and latent difficulty vector $\boldsymbol{\delta} = [\delta_1, ..., \delta_Q]^T$. Specifically,

$$p(x_{nq} = 1 | \boldsymbol{\alpha}, \boldsymbol{\delta}) = \sigma(\alpha_n - \delta_q)$$

where $\sigma$ is the sigmoidal nonlinearity function. He plans to estimate the deterministic latent parameters in the model given the binary data matrix $\boldsymbol{X}$ of dimension $N \times Q$ containing elements $[x_{nq}]$ (assuming that variables $x_{nq}$ are i.i.d.).

   (a) Find the total data likelihood under the given model for the MLSP exam.

   (b) How can Prof. Sam apply gradient descent to estimate the latent ability of students $\alpha_n$ and latent difficulty of questions $\delta_q$ which maximize the total log-likelihood ?

**(Points 10)**

6. **Speech Enhancement** - Let $\mathbf{y}_t, t = 1, ..., T$ denote clean speech signal which is observed as $\mathbf{z}_t = \mathbf{y}_t + \mathbf{v_t}$, where $\mathbf{v_t}$ is the noise. Let $\lambda_s, \lambda_v$ denote the HMM-GMM for clean speech and noise signal respectively. Let $\mathbf{q} = q_1, q_2, ...q_T$ denotes the state sequence of $\lambda_s$ and $\mathbf{l} = l_1, l_2, ...l_T$ denotes the sequence of mixture component index of emission probabilities $p(\mathbf{y}_t|\mathbf{q}_t, \lambda_s)$. Each $q_t \in \{1, ..., N\}$ where $N$ denotes the number of states in $\lambda_s$ and each $l_t \in \{1, ..., M\}$ where $M$ denotes the number of mixture (all the states have the same number of mixture components) in $p(\mathbf{y}_t|\mathbf{q}_t, \lambda_s)$. The speech enhancement task is to estimate the clean signal $\mathbf{y}_t$ by maximizing $p(\mathbf{y}|\mathbf{z})$. Show that this can be achieved by iteratively maximizing $\Phi(\mathbf{y}, \mathbf{y}')$, where

$$\Phi(\mathbf{y}, \mathbf{y}') = \sum_{\mathbf{q},\mathbf{l}} p(\mathbf{q}, \mathbf{l}|\mathbf{y}') \log p(\mathbf{q}, \mathbf{l}, \mathbf{y}|\mathbf{z})$$

Are there any similarities with EM algorithm ? (**Points** 15)

7. **Paired RBM**

A paired RBM is one which has two visible layers $\mathbf{v}^1$, $\mathbf{v}^2$ each with dimension $n_v$ and two hidden layers $\mathbf{h}^1$, $\mathbf{h}^2$ each with dimension $n_h$. Assuming Bernoulli distributions for both visible and hidden units, the energy function of a paired RBM is given by,

$$E[\mathbf{v}^1, \mathbf{v}^2, \mathbf{h}^1, \mathbf{h}^2] = -(\mathbf{h}^1)^T\mathbf{M}\mathbf{h}^2 - (\mathbf{h}^1)^T\mathbf{W}\mathbf{v}^1 - (\mathbf{h}^2)^T\mathbf{W}\mathbf{v}^2 - \mathbf{c}^T\mathbf{v}^1 - \mathbf{c}^T\mathbf{v}^2 - \mathbf{b}^T\mathbf{h}^1 - \mathbf{b}^T\mathbf{h}^2$$

where $\mathbf{M}, \mathbf{W}, \mathbf{b}, \mathbf{c}$ are the parameters of the RBM. The probability density function of the paired RBM is given as,

$$P[\mathbf{v}^1, \mathbf{v}^2, \mathbf{h}^1, \mathbf{h}^2] = \frac{\exp\big(-E[\mathbf{v}^1, \mathbf{v}^2, \mathbf{h}^1, \mathbf{h}^2]\big)}{Z}$$

where $Z$ is a normalization constant. For the paired RBM, show that,

$$P[\mathbf{v}^1, \mathbf{v}^2|\mathbf{h}^1, \mathbf{h}^2] = \prod_{i=1}^{n_v} p(v_i^1|\mathbf{h}^1) \prod_{j=1}^{n_v} p(v_j^2|\mathbf{h}^2)$$

(**Points 5**)