EXAM 2
Thursday, March 31, 2016

- This is a CLOSED BOOK exam. You may use one sheet (front and back) of handwritten notes.
- No calculators are permitted. You need not simplify explicit numerical expressions.
- There are a total of 100 points in the exam. Each problem specifies its point total. Plan your work accordingly.
- You must SHOW YOUR WORK to get full credit.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

Name: ________________________________
Possibly Useful Formulas

Gaussian (Normal) Distribution  A Gaussian is parameterized by $\vec{\mu}$, $\Sigma$, and $D = \dim(\vec{\mu})$ as
$$N(\vec{x}|\vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})}$$

Gaussian Mixture Model (GMM)  A GMM is parameterized by $c_k$, $\vec{\mu}_k$, and $\Sigma_k$ for $1 \leq k \leq K$ as
$$p_X(\vec{x}) = \sum_{k=1}^K c_k N(\vec{x}|\vec{\mu}_k, \Sigma_k)$$

Hidden Markov Model (HMM)  An HMM is parameterized by $\lambda = \{\pi_i, a_{ij}, b_j(\vec{x})\}$, where
$$\pi_i = p(q_1 = i|\lambda), \ 1 \leq i \leq N$$
$$a_{ij} = p(q_{t+1} = j|q_{t} = i, \lambda), \ 1 \leq i, j \leq N$$
$$b_j(\vec{x}) = p(\vec{x}|q_{t} = j, \lambda), \ 1 \leq j \leq N$$

The acoustic model $b_j(\vec{x})$ might be GMM, for example, in which case the HMM parameters include
$$c_{jk} = p(g_t = k|q_{t} = j)$$
$$\vec{\mu}_{jk} = E[\vec{x}_t|q_{t} = j, g_t = k]$$
$$\Sigma_{jk} = E[(\vec{x}_t - \vec{\mu}_{jk})(\vec{x}_t - \vec{\mu}_{jk})^T|q_{t} = j, g_t = k]$$

Scaled Forward Algorithm
$$\hat{\alpha}_1(i) = \pi_i b_i(\vec{x}_1), \ 1 \leq i \leq N$$
$$g_1 = \sum_{i=1}^N \hat{\alpha}_1(i)$$
$$\hat{\alpha}_1(i) = \frac{1}{g_1} \hat{\alpha}_1(i)$$
$$\hat{\alpha}_t(j) = \sum_{i=1}^N \hat{\alpha}_{t-1}(i)a_{ij}b_j(\vec{x}_t)$$
$$g_t = \sum_{j=1}^N \hat{\alpha}_t(j)$$
$$\hat{\alpha}_t(j) = \frac{1}{g_t} \hat{\alpha}_t(j)$$
$$p(\vec{x}_1, \ldots, \vec{x}_t|\lambda) = \prod_{t=1}^T g_t$$
Problem 1  (20 points)

You want to classify zoo animals. Your zoo only has two species: elephants and giraffes. There are more elephants than giraffes: if $Y$ is the species,

\[ p_Y(\text{elephant}) = \frac{e}{e+1} \]
\[ p_Y(\text{giraffe}) = \frac{1}{e+1} \]

where $e = 2.718 \ldots$ is the base of the natural logarithm. The height of giraffes is Gaussian, with mean $\mu_G = 5$ meters and variance $\sigma^2_G = 1$. The height of elephants is also Gaussian, with mean $\mu_E = 3$ and variance $\sigma^2_E = 1$. Under these circumstances, the minimum probability of error classifier is

\[ \hat{y}(x) = \begin{cases} 
\text{giraffe} & x > \theta \\
\text{elephant} & x < \theta 
\end{cases} \]

Find the value of $\theta$ that minimizes the probability of error.
Problem 2  (20 points)

A 3-nearest neighbor (3NN) estimator of $p_{Y|X}(y_0|x_0)$ is computed by finding the 3 nearest neighbors of vector $x_0 = [x_{10}, x_{20}]^T$, then measuring

$$p_{Y|X}(y_0|x_0) = \frac{\# \text{ neighbors from class } y_0}{3}$$

Suppose that the training dataset contains six labeled $(\bar{x}_n, y_n)$ pairs, given by

$$[y_1, y_2, y_3, y_4, y_5, y_6] = [0, 0, 0, 1, 1, 1]$$

$$[\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \bar{x}_6] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ -1 & 0 & -3 & 0 & 3 & 3 \end{bmatrix}$$

Find the 3NN estimator $p_{Y|X}(y_0 = 1| \begin{bmatrix} x_{10} \\ 0 \end{bmatrix})$ as a function of $\bar{x}_0 = \begin{bmatrix} x_{10} \\ 0 \end{bmatrix}$, that is, for $x_{20} = 0$. 


Problem 3  (20 points)

Random vector $X$ is distributed as

$$p_X(\bar{x}) = \sum_{k=1}^{2} c_k \mathcal{N}(\bar{x}|\bar{\mu}, \Sigma)$$

where $c_1 = c_2 = 0.5$, and

$$\bar{\mu}_1 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \bar{\mu}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

Draw a contour plot showing $p_X(\bar{x})$ as a function of $\bar{x}$. Mark the modes of the distribution, and draw contour lines at levels of $e^{-1/2}$ and $e^{-2}$ times the height of the modes.
Problem 4  (20 points)

A particular hidden Markov model is parameterized by $\lambda = \{\pi, a_{ij}, b_j(\tilde{x})\}$ where $\pi_i$ is uniform ($\pi_i = \frac{1}{N}$). Devise an algorithm to compute $p(q_1 = k|\tilde{x}_1, \ldots, \tilde{x}_T, \lambda)$. Your algorithm should be similar to the forward algorithm, but with a different initialization.
Problem 5  (20 points)

The scaled forward algorithm is provided for you on the formula page at the beginning of this exam. In terms of the quantities $\pi_i, a_{ij}, b_j(\bar{x}), \hat{\alpha}_t(j), g_t$, and/or $\hat{\alpha}_t(j)$, find a formula for the quantity $p(q_{t-1} = i, q_t = j, \bar{x}_{t-1}, \bar{x}_t|\bar{x}_1, \ldots, \bar{x}_{t-2}, \lambda)$. 