This is a CLOSED BOOK exam. You may use two pages, both sides, of notes.

There are a total of 100 points in the exam (15-20 points per problem). 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>6</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
</tr>
</tbody>
</table>
Problem 1 (15 points)

Measurements, $x$, are drawn from the pdf

$$p(x) = \Pr\{C = 1\} p_1(x) + \Pr\{C = 2\} p_2(x)$$

$$p_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$p_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}$$

Suppose $\Pr\{C = 1\} = \frac{1}{3}$. Specify the decision rule $y(x)$ that minimizes the probability $\Pr\{y(x) \neq C\}$.

Solution

$$y(x) = \begin{cases} 
1 & x \leq \frac{1}{2} - \ln 2 \\
2 & \text{otherwise} 
\end{cases}$$
Problem 2  (15 points)

Suppose you have $N$ samples $x^{(n)}$, $1 \leq n \leq N$, each distributed i.i.d. as

$$p(x^{(n)}) = \begin{cases} \lambda e^{-\lambda x^{(n)}} & x^{(n)} \geq 0 \\ 0 & x^{(n)} < 0 \end{cases}$$

The parameter $\lambda$ is unknown; in fact, it is itself a random variable, and was selected, prior to creation of this dataset, according to the prior distribution

$$p(\lambda) = \begin{cases} \tau e^{-\tau \lambda} & \lambda \geq 0 \\ 0 & \lambda < 0 \end{cases}$$

Find the MAP estimate of $\lambda$ in terms of $N$, $\tau$, and the samples $x^{(n)}$.

Solution

$$\lambda_{MAP} = \frac{N}{\tau + \sum_{n=1}^{N} x^{(n)}}$$
Problem 3  (20 points)

A two-dimensional real vector \( \vec{x} = [x_1, x_2]^T \) is selected from one of two uniform pdfs, either \( p_1(\vec{x}) \) or \( p_{-1}(\vec{x}) \), given as

\[
p_1(\vec{x}) = \begin{cases} 
\frac{1}{9} & -1 \leq x_2 \leq 2, \quad -\frac{3}{2} \leq x_1 \leq \frac{3}{2} \\
0 & \text{otherwise}
\end{cases}
\]

\[
p_{-1}(\vec{x}) = \begin{cases} 
\frac{1}{9} & -2 \leq x_2 \leq 1, \quad -\frac{3}{2} \leq x_1 \leq \frac{3}{2} \\
0 & \text{otherwise}
\end{cases}
\]

A classifier is trained with the decision rule \( y(\vec{x}) = \text{sign}(\vec{w}^T \vec{x}) \). The weight vector is trained using stochastic gradient descent, with a perceptron training criterion. Let \( \vec{w}^{(n)} \) be the weight vector after presentation of \( \vec{x}^{(n)} \) and \( t^{(n)} \), thus

\[
\vec{w}^{(n)} = \vec{w}^{(n-1)} - \nabla \vec{w} \max \left( 0, -([\vec{w}^{(n-1)}]^T t^{(n)} \vec{x}^{(n)}) \right)
\]

Suppose that after \( N - 1 \) training iterations, for some very large \( N \), the weight vector is given by

\[
\vec{w}^{(N-1)} = \begin{bmatrix} 0 \\ 5000 \end{bmatrix}
\]

Find the expected value after the next iteration, \( E \left[ \vec{w}^{(N)} \bigg| \vec{w}^{(N-1)} \right] = \begin{bmatrix} 0 \\ 5000 \end{bmatrix} \). Be sure to consider the possibility that \( \vec{x}^{(N)} \) might be correctly classified.

Solution

\[
E \left[ \vec{w}^{(N)} \bigg| \vec{w}^{(N-1)} \right] = \begin{bmatrix} 0 \\ 5000 \end{bmatrix} = \Pr \left\{ t^{(N)} = 1 \right\} \Pr \left\{ \text{error} \big| t^{(N)} = 1 \right\} E \left[ \vec{w}^{(n)} \big| \vec{w}^{(n-1)}, t^{(N)} = 1, \text{error} \right] + \Pr \left\{ t^{(N)} = 1 \right\} \Pr \left\{ \text{no error} \big| t^{(N)} = 1 \right\} E \left[ \vec{w}^{(n)} \big| \vec{w}^{(n-1)}, t^{(N)} = 1, \text{no error} \right] + \Pr \left\{ t^{(N)} = -1 \right\} \Pr \left\{ \text{error} \big| t^{(N)} = -1 \right\} E \left[ \vec{w}^{(n)} \big| \vec{w}^{(n-1)}, t^{(N)} = -1, \text{error} \right] + \Pr \left\{ t^{(N)} = -1 \right\} \Pr \left\{ \text{no error} \big| t^{(N)} = -1 \right\} E \left[ \vec{w}^{(n)} \big| \vec{w}^{(n-1)}, t^{(N)} = -1, \text{no error} \right]
\]

\[
= \Pr \left\{ t^{(N)} = 1 \right\} \left( \vec{w}^{(N-1)} - \frac{1}{3} E \left[ \vec{x} \big| t^{(N)} = 1, x_2 < 0 \right] \right) + \Pr \left\{ t^{(N)} = -1 \right\} \left( \vec{w}^{(N-1)} - \frac{1}{3} E \left[ \vec{x} \big| t^{(N)} = -1, x_2 > 0 \right] \right)
\]

\[
= \Pr \left\{ t^{(N)} = 1 \right\} \left( \vec{w}^{(N-1)} + \frac{1}{3} \begin{bmatrix} 0 \\ -0.5 \end{bmatrix} \right) + \Pr \left\{ t^{(N)} = -1 \right\} \left( \vec{w}^{(N-1)} - \frac{1}{3} \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \right)
\]

\[
= \begin{bmatrix} 0 \\ 4999 \frac{5}{6} \end{bmatrix}
\]
Problem 4  (20 points)

A “spiral network” is a brand new category of neural network, invented just for this exam. It is a network with a scalar input variable $x^{(n)}$, a scalar target variable $t^{(n)}$, and with the following architecture:

$$z^{(n)}_j = \begin{cases} x^{(n)} & j = 1 \\ g(a^{(n)}_j) & 2 \leq j \leq M \\ a^{(n)}_j = \sum_{i=1}^{j-1} w_{ji} z^{(n)}_i \end{cases}$$

Suppose that the network is trained to minimize the sum of the per-token squared errors $E^{(n)} = \frac{1}{2} (z^{(n)}_M - t^{(n)})^2$. The error gradient can be written as

$$\frac{\partial E^{(n)}}{\partial w_{ji}} = \delta^{(n)}_j z^{(n)}_i$$

Find a formula that can be used to compute $\delta^{(n)}_j$, for all $2 \leq j \leq M$, in terms of $t^{(n)}$, $z^{(n)}_j = g(a^{(n)}_j)$, and/or $g'(a^{(n)}_j) = \frac{\partial g}{\partial a^{(n)}_j}$.

**Solution**

$$\delta^{(n)}_j = \begin{cases} (z^{(n)}_M - t^{(n)}) g'(a^{(n)}_M) & j = M \\ \sum_{k=j+1}^{M} \delta^{(n)}_k w_{kj} g'(a^{(n)}_j) \end{cases}$$
Problem 5  (15 points)

Exact computation of the Hessian is usually impractical, but there is one case in which it is computationally efficient. Consider a one-layer, one-output network with input \( \vec{x}^{(n)} \in \mathbb{R}^D \) and scalar output \( y^{(n)} \) given by

\[
y^{(n)} = g \left( a^{(n)} \right), \quad a^{(n)} = \sum_{i=1}^{D} w_i x_i^{(n)}
\]

The \((i, j)\)th element of the Hessian matrix is defined by

\[
H(i, j) = \frac{\partial^2 E}{\partial w_i \partial w_j}, \quad E = \frac{1}{2} \sum_{n=1}^{N} \left( y^{(n)} - t^{(n)} \right)^2
\]

Find \( H(i, j) \) exactly in terms of \( w_i, w_j, y^{(n)} = g(a^{(n)}), \) \( g'(a^{(n)}) = \frac{\partial g}{\partial a^{(n)}} \), and \( g''(a^{(n)}) = \frac{\partial^2 g}{\partial a^{(n)} \partial a^{(n)}} \).

Solution

\[
H(i, j) = \sum_{n=1}^{N} x_i^{(n)} x_j^{(n)} \left( (g'(a^{(n)}))^2 + (y^{(n)} - t^{(n)}) g''(a^{(n)}) \right)
\]
Problem 6  (15 points)

Consider an RBM with a scalar real-valued input, \( v \in \mathbb{R} \), and a binary hidden node, \( h \in \{0, 1\} \). Consider the model

\[
p(h, v) = \frac{1}{Z} e^{-E(h,v)}, \quad E(h, v) = \frac{1}{2} (v - (wh + b))^2 - hc
\]

for scalars \( w, b, \) and \( c \) and for denominator

\[
Z = \sum_{h=0}^{1} \int_{-\infty}^{\infty} e^{-E(h,v)} dv
\]

Assume that the values of \( h \) and \( v \) are given; find \( \frac{\partial \ln p(h,v)}{\partial c} \).

Solution

\[
\frac{\partial \ln p(h,v)}{\partial c} = h - \frac{1}{Z} \int_{-\infty}^{\infty} e^{-E(1,v)} dv = h - \frac{1}{1 + e^{-c}}
\]