ME 466 Introduction to AI
Fall 2021
Take-home Final Exam
Due: 19 January 2022 @16:00

The solutions of this exam should be turned in according to the guidelines for programming assignments in the course syllabus. Please read the syllabus first. Submit all evidence of your work; i.e., code, algorithms, flowcharts, MATLAB Command Window/command prompts, inputs, outputs, plots, results, error messages, etc. as an appendix or in the main body of your submission, with proper explanations. Also submit your code and a text file with instructions to run it. **Please fill in the page at the end of this document, sign it, and use it as the cover page of your exam paper. You have to submit this exam in printed form in addition to uploading online.**

MATHEMATICAL BACKGROUND (BONUS SECTION): This part wraps up last week's lecture by summarizing the mathematical background of the Discrete Fourier Transform (DFT). You will get bonus credits if you solve the two simple problems in this section.

1. (15 pts) Prove the following statements:

(a) (5 pts) Let $a \neq 1$ be a complex number. Show that $\sum_{n=0}^{N-1} a^n = \frac{1-a^N}{1-a}$.

(b) (5 pts) Let m be an integer. Show that $e^{j\frac{2\pi m}{N}} = e^{j\frac{2\pi (m \mod N)}{N}}$.

(c) (5 pts) Let m be an integer between 0 and $N-1$. Show that $\sum_{n=0}^{N-1} e^{j\frac{2\pi m}{N}n} = \begin{cases} N & \text{if } m = 0 \\ 0 & \text{if } m = 1, 2, \ldots, N-1 \end{cases}$

Let $\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix}$ be two vectors in N-dimensional complex space $\mathbb{C}^N$. The dot product of these vectors is defined as

$$\mathbf{x} \cdot \mathbf{y} = x_0 y_0^* + x_1 y_1^* + \cdots x_{N-1} y_{N-1}^* = \sum_{n=0}^{N-1} x_n y_n^*$$

where $(\ )^*$ denotes the complex conjugate. Note that the result of the dot product is a complex number. In matrix notation, $\mathbf{x} \cdot \mathbf{y} = \mathbf{y}^H \mathbf{x}$ where $(\ )^H$ is the complex conjugate transpose (or, the Hermitian).

Note that in this case, $\mathbf{x} \cdot \mathbf{y} = (\mathbf{y} \cdot \mathbf{x})^*$

We say that $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** if $\mathbf{x} \cdot \mathbf{y} = 0$. All other definitions and properties related to the dot product for real vectors that you have learned before apply to this dot product definition as well. For example, if $\mathbf{x}$ is real,

$$||\mathbf{x}||^2 = x_0^2 + x_1^2 + \cdots x_{N-1}^2 = \mathbf{x}^T \mathbf{x}\,,$$

but if $\mathbf{x}$ is a complex vector, then

$$||\mathbf{x}||^2 = |x_0|^2 + |x_1|^2 + \cdots |x_{N-1}|^2 = \mathbf{x}^H \mathbf{x}$$

since $zz^* = z^* z = |z|^2$ for a complex number z.

Consider a discrete-time signal of finite length N; the first term is $x[0]$ and the last term is $x[N-1]$.

Then, we can also consider this as a vector in $\mathbb{C}^N$ as $\mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}$.

The Discrete Fourier Transform (DFT) of this signal is another discrete index signal of length N, and is given by

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j\frac{2\pi k}{N}n} \quad , \quad k = 0, 1, \ldots, N-1$$

Note that this is the dot product of $\mathbf{x}$ and the vector $\mathbf{q}_k = \begin{bmatrix} 1 \\ e^{j\frac{2\pi}{N}k} \\ e^{j\frac{2\pi}{N}2k} \\ e^{j\frac{2\pi}{N}3k} \\ \vdots \\ e^{j\frac{2\pi}{N}(N-1)k} \end{bmatrix}$, i.e., $X[k] = \mathbf{q}_k{}^H\mathbf{x}$.

2. (10 pts) Show that $\mathbf{q}_k \cdot \mathbf{q}_\ell = \mathbf{q}_\ell{}^H\, \mathbf{q}_k = \sum_{n=0}^{N-1} e^{j\frac{2\pi k}{N}n} e^{-j\frac{2\pi \ell}{N}n} = \begin{cases} N & \text{if } k = \ell \\ 0 & \text{if } k \neq \ell \end{cases}$, where k and ℓ are two integers between 0 and $N-1$. This means that $\mathbf{q}_k$ and $\mathbf{q}_\ell$ are orthogonal if $k \neq \ell$.

Therefore, $\{\mathbf{q}_0, \ldots, \mathbf{q}_{N-1}\}$ is a set of orthogonal vectors, each with length $\sqrt{N}$, and they form an orthogonal basis for $\mathbb{C}^N$. We can make this basis orthonormal by dividing by the length, and express any vector $\mathbf{x}$ in this orthonormal basis as

$$\mathbf{x} = c_0 \frac{\mathbf{q}_0}{\sqrt{N}} + \cdots + c_{N-1} \frac{\mathbf{q}_{N-1}}{\sqrt{N}}$$

The coefficient c_k is the projection of $\mathbf{x}$ onto the unit vector $\frac{\mathbf{q}_k}{\sqrt{N}}$; i.e., $c_k = \frac{1}{\sqrt{N}} \mathbf{q}_k{}^H \mathbf{x} = \frac{1}{\sqrt{N}} X[k]$.
Therefore,

$$\mathbf{x} = \frac{1}{N}(X[0]\mathbf{q}_0 + \cdots + X[N-1]\mathbf{q}_{N-1}) = \frac{1}{N} \sum_{k=0}^{N-1} X[k]\mathbf{q}_k$$

so the n^{th} element of $\mathbf{x}$ can be obtained by calculating the n^{th} element of the sum above:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi k}{N}n} \quad , \quad n = 0, 1, \ldots, N-1$$

known as the inverse DFT.

END OF BONUS SECTION

Note that the DFT of $x[n]$ is given by

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j\frac{2\pi k}{N}n} \quad , \quad k = 0, 1, \ldots, N-1$$

and the inverse DFT is given by

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi k}{N}n} \quad , \quad n = 0, 1, \ldots, N-1$$

As we discussed in class, $\frac{2\pi k}{N}$ corresponds to $\frac{2\pi f}{f_s}$ where f_s is the sampling frequency of the continuous-time signal. Thus if $|X[k]|$ is large, this means that there is a dominant component in the signal with frequency $f = \frac{k}{N} f_s$. Due to conjugate symmetry, this is unambiguously valid for $\frac{k}{N} \leq \frac{1}{2}$.

The DFT can also be written as a matrix-vector product:

$$\begin{pmatrix} X[0] \\ X[1] \\ \vdots \\ X[N-1] \end{pmatrix} = \underbrace{\begin{pmatrix} \quad\quad\quad\quad\quad\quad \\ \\ \\ \\ \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{pmatrix}$$

where $\mathbf{F}$ is called the DFT matrix and the inverse of $\mathbf{F}$ would be the inverse DFT matrix. You can read the mathematical background in the bonus section and convince yourself that $\mathbf{F}^{-1} = \frac{1}{N}\mathbf{F}^H$. Note that these definitions may be slightly different in different books or resources.

3. (10 pts) Determine the DFT matrices for $N = 2$, $N = 4$, $N = 6$, $N = 8$. (Note that $e^{j\frac{\pi}{2}} = j$, $e^{j\pi} = -1$, etc.)

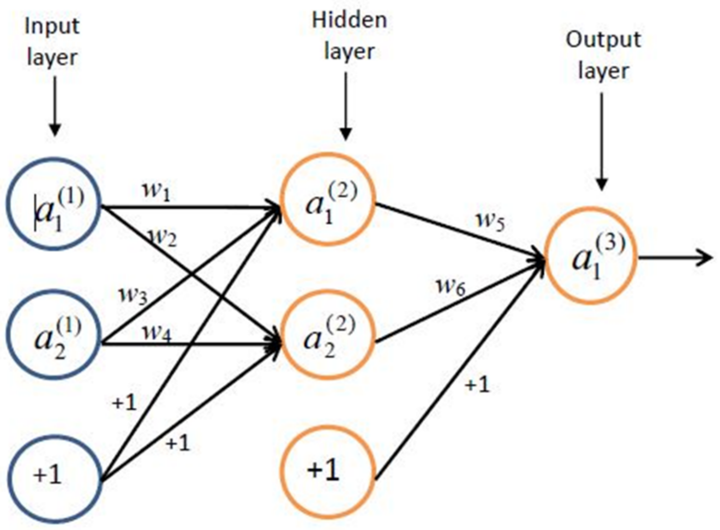
4. (25 pts) Suppose there are only two types of coins in the world, w_0 and w_1. The w_0-type coins have a probability $p = 1/2$ of landing heads, so they are fair coins. w_1-type coins have a probability $p = 1/4$ of landing heads, so they are biased coins. Assume that fair coins are three times more frequent in the world than biased coins. Someone hands you a coin and wants you to tell them if it is fair or biased.

(a) (5 pts) Before making any experiment with the coin, what is your decision? Which category is the coin more likely to belong to? What is your probability of error?

(b) (20 pts) Having taken ME 466, you know that you can make a better decision by making observations, i.e., by simply tossing the coin several times and looking at the outcome. You toss the coin n times, and k out of these n times the coin lands heads. Determine the decision rule that gives the minimum probability of error (as a function of k and n) to decide on the type of the coin.

Hint: The probability of obtaining k heads in n tosses is $\binom{n}{k} p^k (1-p)^{n-k}$, where p is the probability that the coin lands heads in a single toss.

Answer to part (b): "Decide w_1 (biased) if ($k < 0.369\, n - 1$); decide w_0 (fair) otherwise."



5. (20 pts) A neural network is designed with two inputs, one hidden layer with two neurons and one output. In this example, bias values and their weights are chosen to be +1. The hyperbolic tangent function is used for both the hidden layer and the output layer. Let $a_1^{(1)} = 0.6$, $a_2^{(1)} = 0.3$, $w_1 = 0.7$, $w_2 = 0.1$, $w_3 = 0.7$, $w_4 = 0.35$, $w_5 = 0.3$, $w_6 = 0.25$.

(a) (5 pts) Perform the forward pass and determine the output $a_1^{(3)}$.

(b) (15 pts) Let the error be defined as $E = \frac{1}{2}\left(t - a_1^{(3)}\right)^2$ where t is the target output. If $t = 0.7$, calculate $\frac{\partial E}{\partial w_2}$ and determine the updated value of w_2 after the first backpropagation. Let the learning rate $\eta = 1$. You can only use a calculator, MATLAB is not allowed.

6. (20 pts) Consider a neuron with two inputs x_1 and x_2, bias term θ, and output o. Suppose that this is not the ordinary neuron that we have been using, but a "super-neuron" with the following input-output relation:

$$o = w_1x_1 + w_{12}x_1x_2 + w_2x_2 - \theta$$

where w_1, w_{12}, w_2 are weight values like an ordinary neuron. Suppose there is no activation function and o is directly the neuron output.

(a) (10 pts) Let the error be defined as $E = \frac{1}{2}(t - o)^2$ where t is the target output. Derive the gradient descent learning equations for this neuron.

(b) (10 pts) Recall that a single ordinary neuron cannot realize the XOR operation since the patterns are not linearly separable. But this "super-neuron" can. Show (or explain in words, or prove mathematically) how this can be done.
(Hint: Consider how $(w_1x_1 + w_{12}x_1x_2 + w_2x_2 - \theta \lesseqgtr 0)$ regions look like in the x_1–x_2 plane.)

7. (25 pts) The file `finalq7.mat` contains raw data from an accelerometer attached to the leg while a wearable sensing experiment was conducted with 7 different participants. Each row of the matrix `data` corresponds to a 5-second recording of the acceleration data sampled at 25 Hz, while the participant is walking on a straight path. The participant number for each row appears in the variable `participants`, and the sex of the participant is in the variable `gender` in the corresponding row (1 is female, 2 is male). In this problem, the task is to recognize the participant and their gender from the frequency information in their walking data.

The dataset is balanced; each participant has sixty 5-second recordings for walking, resulting in a total of 420 instances. Follow this procedure:

(a) As a rough compensation for the gravitational acceleration, subtract the means from the 5-second samples.

(b) Calculate the DFT of each row. Note that it will be conjugate symmetric. That is, if $X[k]$ denotes the DFT, $X[k] = X^*[125 - k]$ for k between 0 and 124. We will only consider the magnitude information, and $|X[k]| = |X[125 - k]|$, so discard the last 62 samples and use the first 63 samples as features.

(c) Normalize the features in the range [0,1].

(d) From the 420 instances, randomly select 250 instances for training and 50 instances for validation. Use the remaining 120 instances for testing. Build neural networks with the following:

  i. Use the 63 features as input to the network, and a hidden layer with a number of neurons of your choice. Use 7 output neurons, each of which will correspond to one participant, and use the hyperbolic tangent activation function for all neurons. At the output, define the target outputs such that if the correct participant is # 4, the target output should be $[-1 \quad -1 \quad -1 \quad +1 \quad -1 \quad -1 \quad -1]^T$. After each training epoch, use the validation set to calculate the validation error. Stop the training iterations when the validation error starts to increase. Then use the test set as input, decide on the class corresponding to the neuron that gives the maximum output value, and calculate the confusion matrix for the test set.

  ii. Using the 63 features as input, design a neural network to recognize the gender of the walking participant. Use a network architecture of your choice; define the error and the criteria for stopping the training procedure. Again, report the test set confusion matrix.

(e) From the 420 instances, randomly select 300 instances for training. Using the same 63 features, train a k-nearest neighbor classifier for $k = 5$ (In MATLAB, it takes a single line of code to do this if you install the Statistics and Machine Learning Toolbox. Simply use the `fitcknn()` function. There is a similar built-in function in Python as well). Then, test the classifier using the remaining 120 instances. Report the confusion matrix.

ME 466 Introduction to AI
Fall 2021
Take-home Final Exam
Submitted on: 19 January 2022

Name:

Student ID:

Grade:

1	
2	
3	
4	
5	
6	
7	
Σ	

I hereby declare that the paper I am submitting under this cover is product of my own efforts only. Even if I worked on some of the problems together with my classmates, I prepared this paper on my own, without looking at any other classmate's paper. I am knowledgeable about everything that is written under this cover, and I am prepared to explain any scientific/technical content written here if a short oral examination about this paper is conducted by the instructor. I am aware of the serious consequences of cheating.

Signature: ____________________