Name:

|  |  |  |
| --- | --- | --- |
| Problem | Points | Score |
| 1(a) | 10 |  |
| 1(b) | 10 |  |
| 1(c) | 10 |  |
| 1(d) | 20 |  |
| 2(a) | 10 |  |
| 2(b) | 10 |  |
| 3(a) | 20 |  |
| 3(b) | 10 |  |
| Total | 100 |  |

Notes:

1. The exam is closed books and notes. You are allowed on double-sides 8.5x11 in. page of notes. If you don’t have an equation that you think you need, please ask me during the exam.
2. Please clearly indicate your answer to the problem.
3. Note that ungrammatical sentences, incoherent statements, or general illegible scratches will get zero credit.
4. If I can’t read or follow your solution, it is wrong, and no partial credit will be awarded.

**Problem 1:** There are several fundamental concepts in this course. Succinctly explain the concept and discuss how it impacts the machine learning problem.

1. (10) Splitting the data in /train, /dev and /eval.
2. (10) Statistical significance.
3. (10) Sensitivity, specificity and false alarm rate (number of false alarms per unit time).
4. (20) You have developed a great algorithm for segmenting and identifying important events in a signal. Your system detects two states: SEIZ and BCKG. Below is a typical output showing the regions detected as SEIZ:



Explain how you will evaluate the system. Note that the label “SEIZ” occurs only 1% of the time in the data. That poses some very specific challenges when it comes to scoring.

**Problem 2:** Consider the application shown below.



* 1. (10) Design a neural network based system to detect the events shown in black. Note that the events can occur anywhere in the pattern, and the events can be any length (duration). The patterns can be infinitely long and of different lengths (durations). Don’t simply say “I am going to use a transformer”. Explain what your system does and why you think it is appropriate for this problem. Do not simply cite some well-known system (e.g., ChatGPT 4 or Inception XXX). Explain conceptually what your system needs to do to solve the above problem. Hint: IMHO it should have multiple components.
	2. (10) Design a dynamic programming based system to solve this problem. Discuss the strengths and weaknesses of this approach compared to (a).

**Problem No. 3:** For the following equation:

$$o\_{d}=w\_{0}+w\_{1}x\_{1}+\cdots +w\_{n}x\_{n}$$

1. (20) Derive the gradient descent training rule. Define explicitly the cost/error function $E$, assuming that a set of training examples $D$ is provided, where each training example $d \in D$ is associated with the target output $t\_{d}$.
2. (10) Prove that the least mean square error training rule (minimizing the mean square error) performs a gradient descent to minimize the cost/error function $E$ defined in (a).