1. Guidelines for questions types:
   (a) (0 points) Multiple choice, **single** answer.
      A. Mark only **one** the most correct answer.
      B. Mark only one the most correct answer.
   (b) (0 points) Multiple choice, **multiple** answer.
      ◯ Mark **all** the most correct answers.
      ◯ Mark **all** the most correct answers.
   (c) (0 points) Numerical answer

Modeling

1. Which of the following best describes the primary difference between a loss function and an evaluation metric in supervised learning?
   A. A loss function is used during training to compute gradients and update model parameters; an evaluation metric is used after (or during) training to assess model performance on validation/test data.
   B. A loss function is only defined for classification tasks, while an evaluation metric is only defined for regression tasks.
   C. A loss function is minimized during model training, while an evaluation metric is maximized during model training.
   D. A loss function is maximized during model training, while an evaluation metric is minimized during model training.

2. What is correct about gradient descent algorithm:
   √ **A learning rate that is too high may cause parameters to oscillate or even diverge.**
   ◯ Guarantees convergence to a global minimum.
   √ **Choosing a very small learning rate can make GD converge extremely slowly, even though it may still reach a minimum.**
   ◯ Always requires second-order derivatives (the Hessian) to compute its updates.

3. There are two confidence matrices:

		Predicted: Yes	Predicted: No
Actual	Yes	50	10
Actual	No	10	30

		Predicted: Yes	Predicted: No
Actual	Yes	60	0
Actual	No	20	20

   Which metric emphasis the performance difference?
   ◯ Accuracy
   √ **Precision**
   √ **Recall**
   √ **Specificity**

4. The result of logistic regression is $\hat{y}$.

   (a) The interpretation of $\hat{y}$ value is

      A. $\Pr(\hat{y} = 0)$
      B. $\Pr(\hat{y} = 1)$
      C. $\Pr(y = 1)$
      D. $\Pr(\hat{y} = 0|y = 1)$
      E. $\Pr(\hat{y} = 1|y = 1)$
      F. $\Pr(\hat{y} = 0|y = 0)$
      G. $\Pr(\hat{y} = 1|y = 0)$

   (b) The decision is based on $\hat{y} \gtrless \gamma$, where γ is some decision threshold. What are correct statements about γ?

      √ **Larger γ values result in more false negatives.**
      √ **Larger γ values result in more true negatives.**
      ○ Larger γ values result in more false positives.
      ○ Larger γ values result in more true positives.
      ○ Larger γ values result in higher AUC.
      ○ Larger γ values result in lower AUC.

5. (40 points) Given signal $y[n], n = 0, \ldots, L-1$, find μ and a_1 for the biased AR model of the form

$$y[n] = \mu + a_1 y[n-1] \tag{1}$$

You may need the following formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Write only the final steps

Solution:

$$\underbrace{\begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[L-2] \\ x[L-1] \end{bmatrix}}_{\mathbf{y}} = \mu + a_1 \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[L-3] \\ x[L-2] \end{bmatrix} = \underbrace{\begin{bmatrix} \mu & a_1 \end{bmatrix}}_{\mathbf{a}} \underbrace{\begin{bmatrix} 1 & x[0] \\ 1 & x[1] \\ \vdots & \\ 1 & x[L-3] \\ 1 & x[L-2], \end{bmatrix}}_{\mathbf{X}}$$

$$\hat{\mathbf{a}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{y}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} N-1 & \sum_n^{L-2} x[n] \\ \sum_n^{L-2} x[n] & \sum_n^{L-2} x^2[n] \end{bmatrix}$$

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \frac{1}{N\sum_n^{L-2} x^2[n] - \left(\sum_n^{L-2} x[n]\right)^2} \begin{bmatrix} \sum_n^{L-2} x^2[n] & -\sum_n^{L-2} x[n] \\ -\sum_n^{L-2} x[n] & N \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} \sum_{n=1}^{L} x[n] \\ \sum_{n=0}^{L} x[n]x[n-1] \end{bmatrix}$$