Problem 1. (Linear Regression) 2 points

Consider the least squares problem $\min_{w\in\mathbb{R}^{d+1}} \frac{1}{N}\|Xw-y\|_2^2$, with $X \in \mathbb{R}^{N\times(d+1)}$ and $y \in \mathbb{R}^N$. For the following example, is the optimal solution $w^* \in \mathbb{R}^{d+1}$ unique? Justify your answer.

$$X = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 1 & 0 \\ 1 & -3 & 4 \end{bmatrix}, y = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

Problem 2. (Overfitting and underfitting) 4 points

You are addressing a regression problem with $x \in \mathbb{R}^d$ and $y \in \mathbb{R}$. You have tried three different approaches: A, B, C. Each approach gives you a predictor. So, your set of predictors is $\{f_A, f_B, f_C : \mathbb{R}^d \to \mathbb{R}\}$. You obtained the following average train error and test error for each model.

model	train error	test error
A	9.760	9.165
B	0.211	5.072
C	0.633	0.712

1. (1 point) Circle those model(s) that seem to be overfitting: A B C
2. (1 point) Circle those model(s) that seem to be underfitting: A B C
3. (2 points) Circle the model which could profit most from L2 regularization: A B C
   Justify your answer.

Problem 3. (Logistic regression: gradient descent) 4 points

Consider a binary classification problem with data $\{x^i, y^i\}_{i=1}^N$, $x^i \in \mathbb{R}^d$, $y^i \in \{0,1\}$. Let our predictor be 1 if $z^i = w^T x^i + b > 0$ and 0 otherwise. The loss function for training is:

$$L(w,b) = \frac{1}{N}\sum_{i=1}^{N} y^i \log(1+e^{-z^i}) + (1-y^i)\log(1+e^{z^i}).$$

1. (2 points) Write down the gradient of $L(w,b)$ with respect to b i.e. $\frac{\partial L(w,b)}{\partial b}$.

2. (2 points) Complete the gradient descent step below (no need to calculate $\frac{\partial L(w,b)}{\partial w}$).

$$b(t+1) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$w(t+1) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$