Problem 1. ....../6

You are given a dataset containing data on 220 students. Each data point represents a student with her/his hours of study per week (hours), attendance rate (percent), and hours of sleep per night (hours). You want to predict the grade of students.

1. Determine if you are dealing with a regression or a classification problem. Please circle the correct answer.

   (a) For each student the grade is in $\{1, 2, 3, 4, 5, 6\}$: regression classification

   (b) For each student the grade is in $[0, 100]$: regression classification

   Solution:

   (a) For each student the grade is in $\{1, 2, 3, 4, 5, 6\}$: classification

   (b) For each student the grade is in $[0, 100]$: regression

2. Suppose a linear regression is used to predict the grade of students as a function of the features. Specify the dimensions of data set X, and regression parameters w, b.

   Dimensions: $X \in$ ...., $w \in$ ...., $b \in$ ....

   What is the prediction for x^{test}? $f(x^{test}) =$ ....

   Solution:

   $$X \in \mathbb{R}^{220\times 3}, \quad w \in \mathbb{R}^{3\times 1}, \quad b \in \mathbb{R}^{1\times 1}.$$

   What is the prediction for x^{test}? $f(x^{test}) = w^\top x^{test} + b$.

3. You have found that the ratio between study hours (x_1) and sleep hours (x_3) is indicative for student grades. Hence, you define an additional feature:

   $$\phi(x_1, x_3) = \frac{x_1}{x_3}.$$

   (a) Specify the dimensions of the new data set X', and regression parameters w', b':

   $$X' \in \ldots\ldots\ldots, \quad w' \in \ldots\ldots\ldots, \quad b' \in \ldots\ldots\ldots$$

   Solution:

   $$X' \in \mathbb{R}^{220\times 4}, \quad w' \in \mathbb{R}^{4\times 1}, \quad b' \in \mathbb{R}^{1\times 1}.$$

4. After training three different models, you end up with the following training and test errors:

Model	Training Error	Test Error
A	8.76	10.21
B	2.35	9.42
C	1.36	3.79

Match each model (A, B, or C) to the correct description below:

Underfitting: . . . . . . Overfitting: . . . . . . Good generalization: . . . . . .

Solution:

Underfitting: A Overfitting: B Good generalization: C

Problem 2. ....../4

Consider the neural network shown in Figure 1. Here, $x_1, x_2 \in \mathbb{R}$ represent inputs, $\sigma : \mathbb{R} \to \mathbb{R}$ is an arbitrary activation function, and f represents the neural network's output. Weights are written above the respective connections; we assume there are no bias terms.

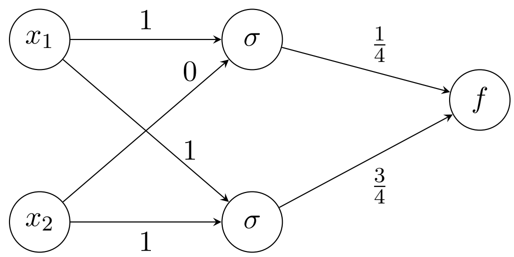

Figure 1

1. Let $\boldsymbol{x} = (x_1, x_2)^\top \in \mathbb{R}^2$, $\boldsymbol{W}_1 \in \mathbb{R}^{2\times 2}$, and $\boldsymbol{w_2} \in \mathbb{R}^2$. Then f can be written in the form

$$f(\boldsymbol{x}) = \boldsymbol{w_2}^\top \boldsymbol{\sigma}(\boldsymbol{W}_1 \boldsymbol{x})$$

where $\boldsymbol{\sigma}(\boldsymbol{x}) = (\sigma(x_1), \sigma(x_2))^\top$. Specify the parameters such that the neural network matches the one in Figure 1:

$$\boldsymbol{W}_1 = \begin{pmatrix} \cdots & \cdots \\ \cdots & \cdots \end{pmatrix}, \qquad \boldsymbol{w}_2 = \begin{pmatrix} \cdots \\ \cdots \end{pmatrix}.$$

Solution: $\boldsymbol{W}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\boldsymbol{w}_2 = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}$.

2. The following four functions may be used as activations in neural networks:

(a) sigmoid: $\sigma(x) = \frac{1}{1+e^{-x}}$

(b) tanh: $\sigma(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

(c) ReLU: $\sigma(x) = \max(0, x)$

(d) Leaky ReLU: $\sigma(x) = \max(0.1x, x)$

Below, we list different properties of activation functions. For each, write the function among (a)-(d) for which this property holds. (Each property holds for exactly one of the functions.)

(1) $f(x) \longrightarrow -1$ as $x \to -\infty$: ....

(2) There exists $x < 0$ such that the derivative $\sigma'(x) = 0$: ....

(3) For all $x \in \mathbb{R}$, $0 \leq \sigma(x) \leq 1$: ....

(4) On the interval $(-\infty, 0)$, $\sigma(x)$ is linear with positive slope: ....

Solution: (1) tanh, (2) ReLU, (3) sigmoid, (4) Leaky ReLU.

3. Evaluate the neural network at the point $x = (1, -2)^\top$ when using a ReLU activation.

$$f(\begin{pmatrix} 1 \\ -2 \end{pmatrix}) = \dots\dots\dots\dots$$

Solution:

$$\begin{aligned}
f\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right) &= \boldsymbol{w}_2^\top \boldsymbol{\sigma}(\boldsymbol{W}_1 \boldsymbol{x}) \\
&= \boldsymbol{w}_2^\top \boldsymbol{\sigma}\left(\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right) \\
&= \boldsymbol{w}_2^\top \begin{pmatrix} \sigma(1) \\ \sigma(-1) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{(since ReLU}(1) = 1 \text{ and ReLU}(-1) = 0) \\
&= \frac{1}{4}.
\end{aligned}$$