Name: _HUANG JIAHUI_Student ID:_44251017_

11040A Neural Networks

Homework 1

Due Date: May 1, 2025 24:00

1. Let $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 3 & 6 \\ 7 & 8 \end{pmatrix}$. Compute AB and BA.

$$AB = \begin{pmatrix} 1\times1+0\times3+1\times7 & 1\times2+0\times6+1\times8 \\ 1\times1+2\times3+3\times7 & 1\times2+2\times6+3\times8 \end{pmatrix} = \begin{pmatrix} 8 & 10 \\ 28 & 38 \end{pmatrix}$$

$$BA = \begin{pmatrix} 1\times1+2\times1 & 1\times0+2\times2 & 1\times1+2\times3 \\ 3\times1+6\times1 & 3\times0+6\times2 & 3\times1+6\times3 \\ 7\times1+8\times1 & 7\times0+8\times2 & 7\times1+8\times3 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 7 \\ 9 & 12 & 21 \\ 15 & 16 & 31 \end{pmatrix}$$

2. Let $\boldsymbol{x}$ be a column vector, $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}$. Compute $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$.

$$f(x) = \frac{1}{2}(x_1^2 + x_2^2 + \cdots + x_n^2)$$

$$\text{then } \nabla_x f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right)^T$$

$$= (x_1, x_2, \cdots, x_n)^T = x$$

3. What are the conditions for a function to have a gradient vector at a point?

The function is differenviable at the point.

All partial derivatives exist and continuous around the point.

Name: _HUANG JIAHUI_Student ID:_44251017_

4. Let $u = \ln(x + y^3 + z^5)$. Compute the gradient at the point $M(1, 3, -2)$.

$\frac{\partial u}{\partial x} = \frac{1}{x+y^3+z^5}$, $\frac{\partial u}{\partial y} = \frac{3y^2}{x+y^3+z^5}$, $\frac{\partial u}{\partial z} = \frac{5z^4}{x+y^3+z^5}$

thus, $\nabla u = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right)$

then $\nabla u\Big|_{(1,3,-2)} = \left(-\frac{1}{4}, -\frac{27}{4}, -20\right)$

5. What is the geometric relationship between a gradient vector and the function?

(hint: the gradient is the direction that the … / the gradient is perpendicular to …)

The gradient is the direction of the steepest ascent and is perpendicular to the level surface.

6. Let $z = e^u \cos v$, $u = xy$, $v = 2x + y^2$. Compute $\frac{\partial z}{\partial x}$, and $\frac{\partial z}{\partial y}$.

$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial z}{\partial v} \cdot \frac{\partial v}{\partial x} = y \cdot e^u \cos v - 2e^u \sin v$

$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u} \cdot \frac{\partial u}{\partial y} + \frac{\partial z}{\partial v} \cdot \frac{\partial v}{\partial y} = xe^u \cos v - 2ye^u \sin v$

7. Let $z = f(x^2y^3,\ xy)$. Compute $\frac{\partial z}{\partial x}$, and $\frac{\partial z}{\partial y}$.

let $u = x^2y^3$, $v = xy$. then $z = f(u, v)$

$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial x} = 2xy^3 \cdot \frac{\partial f}{\partial u} + y \cdot \frac{\partial f}{\partial v}$

$\frac{\partial z}{\partial y} = \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial y} + \frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial y} = 3x^2y^2 \cdot \frac{\partial f}{\partial u} + x \cdot \frac{\partial f}{\partial v}$

Name: _HUANG JIAHUI_Student ID:_44251017_

8. Compute the extremum of the function $f(x,y) = x^3 + y^3 - 3xy$.

$f'_x(x,y) = 3x^2 - 3y$. $f'_y(x,y) = 3y^2 - 3x$

solve $\begin{cases} 3x^2-3y=0 \\ 3y^2-3x=0 \end{cases} \Rightarrow \begin{cases} x^2=y \\ y^2=x \end{cases} \Rightarrow \begin{cases} x=0 \\ y=0 \end{cases}$ or $\begin{cases} x=1 \\ y=1 \end{cases}$

$f(0,0)=0$, $f(1,1)=-1$

thus, $(1,1)$ is a minimum.

9. Suppose there are a set of points $X = \{\boldsymbol{x}_i \in \mathbb{R}^d | i = 1, 2, \ldots, n\}$ and a random point $\boldsymbol{p} \in \mathbb{R}^d$ in d-dimentional space.

Let the Mean Square Error (MSE): MSE $= \frac{1}{n}\sum_i (d(\boldsymbol{x}_i, \boldsymbol{p}))^2$, where $d(\boldsymbol{x}, \boldsymbol{y})$ is the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$: $\mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$

Prove that the MSE is minimized when $\boldsymbol{p}$ is at the centroid ($\boldsymbol{c} = \frac{1}{n}\sum_i \boldsymbol{x}_i$) of the X.

$$MSE(p) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d}(x_{ij} - p_j)^2$$

$$\frac{\partial MSE}{\partial p_j} = \frac{1}{n}\sum_{i=1}^{n} 2(p_j - x_{ij}) = \frac{2}{n}\left(np_j - \sum_{i=1}^{n} x_{ij}\right)$$

let $\frac{\partial MSE}{\partial p_j} = 0 \Rightarrow np_j = \sum_{i=1}^{n} x_{ij} \Rightarrow p_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$

thus MSE is minimized when p is at the centroid of X.

Name: _HUANG JIAHUI_Student ID:_44251017_

10. Explain the following terms (within 30 words respectively):
    (a) supervised learning
    (b) unsupervised learning
    (c) TP, FP, TN, FN, accuracy, precision, recall and f-score
    (d) underfitting and overfitting

(a) supervised learning: Learning from labeled data where correct outputs are known.

(b) unsupervised learning: Learning patterns from data without labels.

(c) TP, FP, TN, FN, Accuracy, Precision, Recall, F1-score:

True/False Positives/Negatives; Accuracy = correct predictions rate; Precision = TP/(TP+FP); Recall = TP/(TP+FN); F1-score = harmonic mean of Precision and Recall.

(d) underfitting and Overfitting:

Underfitting happens when the model is too simple to learn the data patterns.

Overfitting happens when the model is too complex and learns noise along with the data and performs well on training dataset but performs poor on test dataset.

11. Explain the following terms
    (a) bias–variance dilemma
    (b) No Free Lunch Theorem

(a) It describes the trade-off between a model's ability to fit training data (low bias) and its ability to generalize to unseen data (low variance).

(b) No machine learning algorithm performs best for all problems; algorithm performance depends on the specific problem.

12. What is the difference between validation set and test set?

Validation set is used during training to tune model hyperparameters and prevent overfitting.

Test set is used after training to evaluate the final model’s generalization performance.

13. Please name three main challenges in machine learning.
    (a) Overfitting: Overfitting happens when the model is too complex and learns noise along with the data and performs well on training dataset but performs poor on test dataset.
    (b) Data bias: training data may have hidden biases, leading to unfair or skewed models.
    (c) Generalization: models often perform poorly on unseen data even if they fit the training data well.

Programming part

1. Get started with a programming language and learn to run matrix operations using your programming language.
   (You can submit your programming files. Or you can organize your program into a markdown file or a Jupiter Notebook file as a report and submit it.)

   We recommend using the NumPy library in Python, as the TA will use NumPy as an example to explain the programming part. Also, if you plan to use deep learning for research in the future, Python is also a mainstream programming language in academia.

Please pack your files using **zip** and name it in the format of "studentID_name_NN_assignmentID.zip" (e.g. 44278212_Archimedes_NN_1.zip), and Submit it to TA (Xijian RUI): xijian.rui@fuji.waseda.jp