CS 412 Machine Problem 1

1 Question Answering

Q1. What is the difference between supervised learning and unsupervised learning? (10 points)

Q2. What is the difference between regression and classification? (10 points)

Q3. Please list three applications of regression and three applications of classification. Please specify the input and output of each application. An example is given below. (10 points)

Application	Problem	Input	Output
Housing price prediction	Regression	Size of house, score of school district, build year, etc	Housing price
…	…	…	…

Q4. Given two vectors $\mathbf{a} = [1, -1, 0]$ and $\mathbf{b} = [0, 1, 1]$, please manually calculate both the L_1 Norm and L_2 Norm of $\mathbf{a}$, $\boldsymbol{b}$, and $\mathbf{a} - \mathbf{b}$. (10 points)

2 Programming

The goal of this programming assignment is let you be familiar with Python and Numpy and generate data for future use. It is mandatory to use Python 3.

P1. Given two matrices below, please use Numpy to calculate their matrix multiplication and element-wise multiplication. (10 points)

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 1 & 0 & 8 \end{bmatrix} \mathbf{B} = \begin{bmatrix} 7 & 9 & 2 \\ 3 & 4 & 3 \\ 6 & 1 & 3 \end{bmatrix}$$

P2. Please use Numpy to calculate $\mathbf{A^{-1}}$, $\mathbf{A^T}$, $\mathbf{A^T A}$. (10 points)

P3. Please use Numpy to calculate both the L_1 Norm and L_2 Norm of $\mathbf{a} = [1, -1, 0], \mathbf{b} = [0, 1, 1]$, $\mathbf{a} - \mathbf{b}$. (10 points)

P4. Please follow the steps below to generate synthetic data that can be used for regression. (15 points)

1. Use numpy.random.uniform[1] to generate 20 random numbers following a uniform distribution in the range [0,1]. They are samples of a scalar input x.
2. For each sample value of x, generate the ground truth output value $y = \sin(2\pi x) + n$, where n is a random noise value from a Gaussian distribution[2] with 0 mean and 0.3 standard deviation.

[1] https://numpy.org/doc/stable/reference/random/generated/numpy.random.uniform.html
[2] https://numpy.org/doc/stable/reference/random/generated/numpy.random.normal.html

3. Now you have 20 pairs of (x, y). Use matplotlib.pyplot[3] to draw these points. Do not connect points with lines.
4. Your answer to this question is the figure.

P5. Please follow the steps below to generate synthetic data that can be used for classification and clustering. (15 points)

1. Use numpy.random.multivariate_normal[4] to generate 10 2D points from a multivariate Gaussian distribution whose mean is [1, 1] and covariance matrix is a 2-by-2 identity matrix. The label of each point is 0.
2. Similarly, generate 10 2D points from a multivariate Gaussian distribution with mean [1, 2] and an identity covariance matrix. The label of each point is 1.
3. Similarly, generate 10 2D points from a multivariate Gaussian distribution with mean [2, 1] and an identity covariance matrix. The label of each point is 2.
4. Now you have 30 2D points belonging to 3 categories. Use matplotlib.pyplot to draw these points. You should choose 3 different colors for 3 categories and draw each point using the color corresponding to its category. Do not connect points with lines.
5. Your answer to this question is the figure.

3 Optional Programming

O1. Use numpy.linalg.eig[5] to perform eigendecomposition for matrix $\boldsymbol{C}$. Print the matrix of eigen vectors and the diagonal matrix of eigen values. Then, verify whether matrix $\boldsymbol{C}$ can be reconstructed via the matrix of eigen vectors and the diagonal matrix of eigen values. (2.5 points)

$$\mathbf{C} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 8 \end{bmatrix}$$

O2. Use numpy.linalg.svd[6] to perform singular value decomposition for matrix $\boldsymbol{C}$. Print the three obtained matrices. (2.5 points)

4 Submission

Please follow the instructions below for submission.

- You need to upload two files to Blackboard: a PDF file and a .py file[7]. Do not compress them into a single ZIP file.
- The PDF file contains all your solutions to this homework. For Question Answering, you can either type answers or handwrite them and take a photo. For Programming, you need to print the results or draw figures, and take screenshots.

[3] https://matplotlib.org/stable/tutorials/introductory/pyplot.html
[4] https://numpy.org/doc/stable/reference/random/generated/numpy.random.multivariate_normal.html
[5] https://numpy.org/doc/stable/reference/generated/numpy.linalg.eig.html
[6] https://numpy.org/doc/stable/reference/generated/numpy.linalg.svd.html
[7] Using Jupyter Notebook and submitting a .ipynb file instead of a .py file are fine.

- The .py file contains all your code for the programming problems.