THE UNIVERSITY OF HONGKONG
DEPARTMENT OF PHYSICS

Final Exam

Course: *Machine Learning in Physics (PHYS3151)* – Professor: *Dr. Ziyang Meng*
Date: *Dec. 15th, 2020* Time: *9:30 am - 11:30 am*

A: Answer all the questions. Please submit a pdf file containing all the solutions and all your codes used for generating your answers.

B: This is an online open-book exam. You may use computers and all the teaching materials for the exam.

C: Discussing with others during the exam is strictly prohibited.

1. **Multivariate Linear Regression. (25)**

   (a) What is the hypothesis and the cost function of Multivariate Linear Regression?

   (b) Please explain how Gradient Descent method works?

   (c) Quadratic potential is very common in physics systems, for example, the potential energy of a simple harmonic oscillator is quadratic in x (the distance away from the equilibrium position), and therefore physicists are very often asked to perform optimization or finding the minimal of quadratic potentials. Here we have the potential of the form $f(x) = \frac{1}{2}x^T Ax - b^T x + c$, where $A = \begin{pmatrix} 54 & 19 & 26 \\ 19 & 30 & 14 \\ 26 & 14 & 19 \end{pmatrix}$, $b = \begin{pmatrix} 11 \\ 1 \\ 3 \end{pmatrix}$, and $c = 19$. Please use the Steepest Descent algorithm to find the minimum of the potential. Note that you need to write codes from scratch for this problem.

2. **Logistic Regression and SVM. (30)**
   Classifying states of matter is the task of physics, such as classification of liquid, gas and solid, separating normal metal and superconducting metal. Logistic regression and SVM are being used by physicists to solve the classification problems with the data generated from experiments or numerical simulations. Solving the following problem will bestow you with the ability of classify physical problems.

   (a) Please write down the hypothesis and cost function of Logistic Regression and SVM separately and make plots of the cost functions of the two methods as a function of $z = \theta^T x$. What is the relation between the two cost functions? (Hint: discuss the limits when $|z| = \infty$ or 0).

   (b) Please use the two features **SepalLengthCm** and **PetalWidthCm** of Iris.csv to separate the two species **Iris-setosa** and **Iris-versicolor** (draw a decision boundary). Note that you can either use logistical regression or SVM for this problem.

3. **PCA and Clustering. (30)**
   In physics, obtaining eigenvalue and eigenvectors of matrices is a very common practice, for example, in quantum mechnics, one is usually asked to perform singular value decomposition of a Hamiltonian matrix or a wavefuction. Such practice, has roots in the machine learning, in particular in PCA.

(a) Here we have the data matrix

$$X = \begin{pmatrix} 7 & 4 & 3 \\ 4 & 1 & 8 \\ 6 & 3 & 5 \\ 8 & 5 & 7 \\ 8 & 6 & 1 \\ 7 & 2 & 9 \\ 5 & 3 & 3 \\ 9 & 8 & 5 \\ 8 & 2 & 2 \\ 2 & 1 & 1 \end{pmatrix}$$

Please compute the covariance matrix Σ, perform Singular Value Decomposition on Σ, and write down all the eigenvalues and eigenvectors of Σ.

(b) Again, use the data Iris.csv, imagine data **PetalWidthCm** and **Species** in the last two column are damaged. Please use the remaining three features two group the Iris into three clusters using the K-means algorithm and make a 3D plot to represent the clusters.

4. **Neural Networks. (15)**
In the course project, we have solved a 2D Ising model with Metropolis algorithm, and performed logistic regression to find the continuous phase transition point. However, in the modern application of machine learning in Physics, physicists start to employ neural network to detect the different phases, for example the disordered and ordered states in the Ising model setting. It turns out that the AI is capable of finding the order parameter automatically. In this simple question, we will not be able to solve the Ising model with neural networks, however, one is required to show how to optimize the weights in a neural network with back propagation.

(a) Please draw a schematic plot of perceptron and mark all the essential components.

(b) As for the case in Figure 1, if we only have two data: $x^{(1)} = [1, 2]^T$, $y^{(1)} = 1$ and $x^{(2)} = [3, 4]^T$, $y^{(2)} = 2$ and initially we set $\Theta_{k,i}^{(l)} = 1$ for all the neurons. Please use the back propagation algorithm to calculate the partial derivative of the cost function (consider the case without regularization terms) with respect to all $\Theta_{k,i}^{(l)}$ for just one step. Note that the activation function for all the neurons are sigmoid functions.

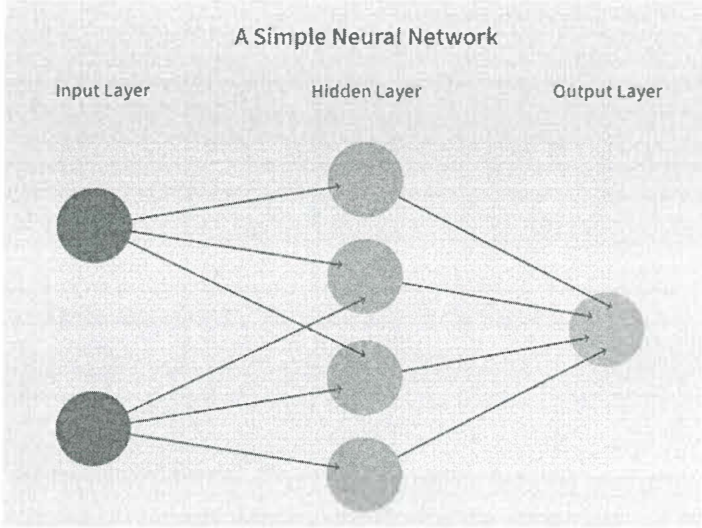

Figure 1: Fully connected feedforward network with one hidden layer and one output layer.

********** END OF PAPER **********