B455: Principles of Machine Learning

HW2 (Due: **Feb. 15 Friday 5pm**)

https://iu.instructure.com/courses/1771430

1. Modify the perceptron implementation (see the notebook pcn_example.ipynb shared with you on the google drive under week 3) to use the adaline model (using the linear activation function). You should modify the code to use the linear activation function and the update algorithm accordingly. You should train your model using the input data for the OR gate (Note: here you can use the real output 0 or 1 to update the weights, while the predicted output will be a real number. You also need to determine a threshold to convert the output into the final output 0 or 1). You should submit your modified implementation in a notebook file through canvas, which should include the implementation as well as a short description of result.

2. Using the linear regression function in NumPy (see page 66 of the ML book) to solve the following problem. You should submit your code in a notebook file through canvas, which should include the implementation and the input data as well as a short discussion about your result. The table below lists the top ten colleges based on mid-career salary and the associated yearly tuition costs. Use the data to check if the higher cost of tuition translate into higher-paying jobs.

School	Mid-Career Salary (in thousands)	Yearly Tuition
Princeton	137	28,540
Harvey Mudd	135	40,133
CalTech	127	39,900
US Naval Academy	122	0
West Point	120	0
MIT	118	42,050
Lehigh University	118	43,220
NYU-Poly	117	39,565
Babson College	117	40,400
Stanford	114	54,506

3. A logistic model can be used to evaluate the effect of a (binary) factor combined with other variables on the observed consequences. Consider the example as follows. In a five-year follow-up study on N disease-free human subjects, researchers aim to assess the effect of the environmental exposure to a heavy metal (E=1, exposed or 0 not exposed) on the development (or not) of a certain disease. The continuous variables of interests are age (AGE) and obesity status (OBS), the environmental factor E of each subject was determined at the start of the study. Devise the logit form of a logistic regression model that assesses the interaction effects of AGE with E and OBS with E, and explain how the model can be learned from the data (where each data point represents a human subject with the known disease status (0/1), AGE, OBS and E). (Hint: the interaction effects of AGE and E can be modeled by a created variable AGEE=AGE * E; similar for OBSE=OBE * E. You also need to consider the effects of the independent variables AGE, OBS and E, respectively).

4. (Problem 4.2) Suppose that the local power company wants to predict electricity demand for the next 5 days. They have the data about daily demand for the last 5 years. Typically, the demand will be a number between 80 and 400.

a) Describe how you could use an MLP to make the prediction. What parameters would you have to choose, and what do you think would be sensible values for them?

b) If the weather forecast for the next day, being the estimated temperatures for daytime and nighttime, was available, how would you add that into your system?

c) Do you think that this system would work well for predicting power consumption? Are there demands that it would not be able to predict?

5. (Problem 4.9) A hospital manager wants to predict how many beds will be needed in the geriatric ward. He asks you to design a neural network method for making this prediction. He has data for the last 5 years that cover:
   - The number of people in the geriatric ward each week.
   - The weather (average day and night temperatures).
   - The season of the year (spring, summer, autumn, winter).
   - Whether or not there was an epidemic on (use a binary variable: yes or no).

Design a suitable MLP for this problem, considering how you would choose the number of hidden neurons, the inputs (and whether there are any other inputs you need) and the preprocessing, and whether or not you would expect the system to work.

6. In the k-nearest neighbor (KNN) method, the prediction can be made by the majority vote when these neighbors do not have the same property (e.g., see below for predicting the color of a new data point U based on its top k=3 nearest neighbors in the space of two features width and height). 1) if all neighbors are given the same weight in the KNN (k=3) method, what will be the prediction? 2) if the vote is made according to the inverse square of distant $w \sim 1/d^2$, what will be the prediction? Justify your answers.