Assignment 2

Deadline: 23:55, November 19, 2023

1. Given the model with parameter θ, the maximum likelihood estimation w.r.t θ can be formulated as:

$$\theta = \arg\max_{\theta} \prod_{n=1}^{N} P(y^{(n)} | \boldsymbol{x}^{(n)}; \theta)$$

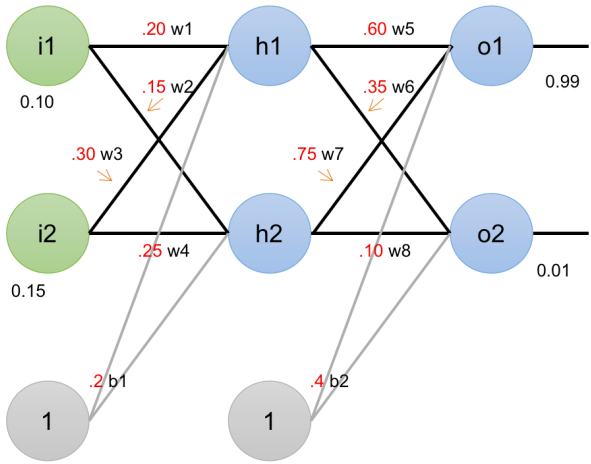
We can use logistic regression to do binary classification, since the output $h_\theta(\boldsymbol{x}) = \frac{1}{1+e^{-f_\theta(\boldsymbol{x})}}$ can be viewed as the probability that the input $\boldsymbol{x}$ belongs to the class 0. Given the objective function in the form of cross entropy loss L_θ :

$$L_\theta = \frac{1}{N} \sum_{n=1}^{N} [-y^{(n)} \ln (1 - h_\theta(\boldsymbol{x}^{(n)})) - (1 - y^{(n)}) \ln (h_\theta(\boldsymbol{x}^{(n)}))]$$

(1) Illustrate the step-by-step process of obtaining L_θ using maximum likelihood estimation. **Show your steps clearly.** (10 marks)

(2) Compute the derivate of L_θ with respect to θ. **Show your steps clearly**. (10 marks)

2. There is a neural network with two inputs, two hidden neurons, two output neurons. Additionally, the hidden and output neurons will include a bias. Backpropagation is a common method for training a neural network. The goal of backpropagation is to optimize the weights so that the neural network can learn how to correctly map arbitrary inputs to outputs. The goal is: given inputs 0.1 and 0.15, the expected outputs of the neural network are 0.99 and 0.01. The initial values for the network parameters are shown in the following figure.



(1) Try to calculate two outputs, i.e., out_{o1} and out_{o2} in the first forward pass step. **Show your steps clearly.** (10 marks)
Hint: forward pass step includes *inputs to hidden layer neurons, squash the total input using an activation function (here we use the sigmoid function)* and *hidden layer outputs to output layer neurons.*

(2) Try to calculate the error for each output neuron using the squared error function and sum them to get the total error E_{total}. **Show your steps clearly**. (10 marks)

(3) Consider w_5. We want to know how to change w_5 in our optimization, a.k.a. $\frac{\partial E_{total}}{\partial w_5}$, the gradient with respect to w_5. Try to calculate the new weight w_5^{new} in the spirit of the gradient descent method with learning rate 0.1. **Show your steps clearly**. (20 marks)

3. You are given the following data set, where

• "name" is not an attribute;

• attributes "eye-color" and "hair-color" are discrete;

• attributes "height" and "weight" are numeric;

• the class label "region" has three possible values: Asia, Europe, and America.

name	height	weight(kg)	eye-color	hair-color	region
Abel	182	73	hazel	black	Europe
Bob	183	68	hazel	blond	Europe
Carl	177	70	blue	brown	Europe
Dale	171	61	blue	brown	Asia
Eric	168	67	blue	blond	Asia
Felix	165	63	brown	black	Asia
George	180	80	brown	brown	America
Howard	185	75	hazel	black	America
Igor	173	73	brown	blond	America

Find the feature that should be used in the decision tree's root according to the information gain. **Show your steps clearly**. (40 marks)