SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

50.007 Machine Learning, Spring 2023
Homework 3

Due Friday 31 March 2023, 11.59pm

In this homework, we would like to look at Logistic Regression, Neural Network, and Naïve Bayes.

1 Logistic Regression [25 points]

Consider 13 data points from a 2-d space where each point is of the form $x = (x_1, x_2)$, as shown in Figure 1. Now we want to train a logistic regression classifier based on the given data. Suppose the hypothesis function of the logistic regression is $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$ where $g(z)$ is the logistic function, and parameter vector $\theta = [\theta_0, \theta_1, \theta_2]^T$ is initialized as $[0, -1, 1]^T$.

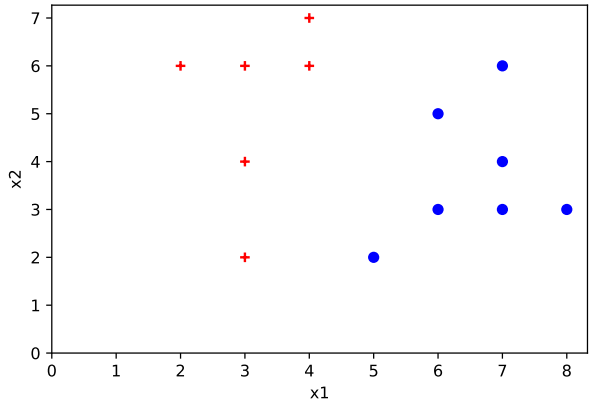

Figure 1: The given toy dataset contains 13 data points, where the red cross indicates that the point belongs to the **positive** class ($y = 1$), and the blue dot indicates that the point belongs to the **negative** class ($y = 0$).

Please write a function implementing gradient descent (simultaneously updating all the parameters, *i.e.*, $\theta_0, \theta_1, \theta_2$) to find suitable parameters for the hypothesis function. We set the learning rate as $\alpha = 0.1$ and run the gradient descent for 150 iterations. Plot the training loss over the 150 iterations using a line plot and write down the final decision boundary.

2 Neural Networks [40 points]

As shown in Figure 2, we are given a 3-layer neural network. Instead of incorporating the bias term into the weight matrix Θ_i, we explicitly write the bias term, resulting in $\mathbf{z}_2 = \mathbf{w}_2\mathbf{a}_1 + b_2$, where $\mathbf{w}_2 \in \mathbb{R}^{1\times 3}$ is the weight vector and $b_2 \in \mathbb{R}$ is the bias term for the mapping function from the hidden layer to the output layer, respectively. The output of this neural network $\hat{y} = \sigma(\mathbf{z}_2)$.

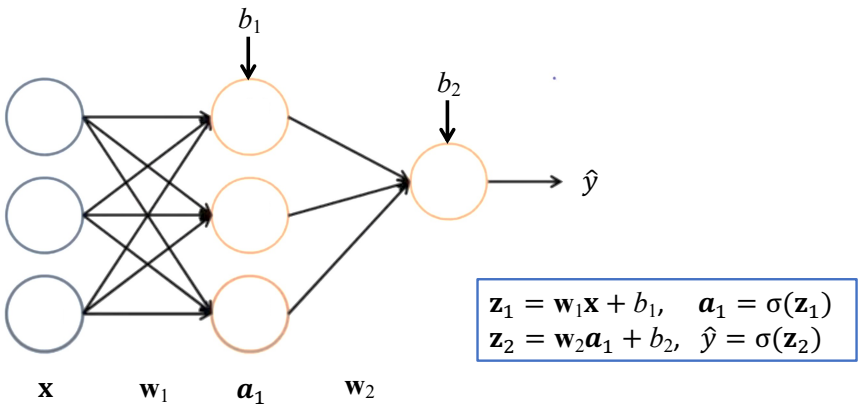

Figure 2: The architecture of a 3-layer neural network with the mathematical representation of forward propagation.

(1) [5 points] Write down the mathematical representation of $\frac{\partial \hat{y}}{\partial \mathbf{w}_2}$ and $\frac{\partial \hat{y}}{\partial b_2}$.

(2) [5 points] Let us assume the activation function σ is the softplus function, please derive the closed-form expression for calculating $\frac{\partial \hat{y}}{\partial \mathbf{w}_2}$ and $\frac{\partial \hat{y}}{\partial b_2}$.

(3) [15 points] If we still use softplus as the activation function but change the value of the bias b_2, does $\frac{\partial \hat{y}}{\partial \mathbf{x}}$ ($\mathbf{x}$ is the input to this 3-layer neural network) change? Please prove your answer. [Hint: the change of bias can be expressed as Δb_2]

(4) [15 points] Now let us assume that the activation function σ is the logistic function, please derive the closed-form expression for calculating $\frac{\partial \hat{y}}{\partial \mathbf{w}_1}$.

3 Naïve Bayes [35 points]

Suppose we have a dataset of individuals who have been audited by the IRS for tax evasion. The data includes three features/attributes: 'Refund' (yes or no), 'Marital Status' (single, married, divorced), and 'Taxable Income' (a continuous value). The target variable is whether or not the individual was found guilty of tax evasion (yes or no). The dataset is shown in the Table 1.

Note that for continuous attribute 'Taxable Income', we assume it follows a class-conditional normal distribution, which means $P(\text{Taxable Income}|\text{Evade Tax} = c) \sim \mathcal{N}(\mu_c, \sigma_c^2)$ where $c \in \{\text{Yes}, \text{No}\}$ is the value of output variable, *i.e.* 'Evade Tax'. Specifically, the probability density function of $\mathcal{N}(\mu_c, \sigma_c^2)$ is $P(x|Y = c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(x-\mu_c)^2}{2\sigma_c^2}\right)$. The sample mean μ_c is computed as $\mu_c = \frac{1}{n_c}\sum_{i=1}^{n_c} x_i$, where n_c is the number of samples *w.r.t.* class c in the training set. The sample variance σ_c^2 is computed as $\sigma_c^2 = \frac{1}{n_c - 1}\sum_{i=1}^{n_c}(x_i - \mu_c)^2$, where having $(n_c - 1)$ instead of n_c in the denominator is because of the use of Bessel's correction.

Your task is to implement a Naive Bayes classifier to predict whether an individual is likely to evade taxes or not, based on his/her refund status (Yes), marital status (Married), and taxable income (79K). Please clearly present the steps that lead to your final predictions.

Refund	Marital Status	Taxable Income	Evade Tax
Yes	Single	125K	No
No	Married	100K	No
No	Single	70K	No
Yes	Married	120K	No
No	Divorced	95K	Yes
No	Married	60K	No
Yes	Divorced	220K	No
No	Single	85K	Yes
No	Married	75K	No
No	Single	90K	Yes

Table 1: The toy dataset of individuals who have been audited by the IRS for tax evasion.