Logistic Regression for Predicting Customer Behavior

Read the simmons.csv file in R which contains data about the Simmons stores. The
simmons.csv dataset contains the following information:

- UsesCoupon: Whether a customer uses a coupon provided by the
- stores.HasCard: Whether the customer has a store card or not.
- Spends: The total amount spent by the customer in the last year in the unit of
  $1000.
- Customer: The id of the customer.

The aim of the following exercise is to predict the probability that customers will use the coupons provided to them by a store (predicting the coupon variable below based on the spending level and whether the customer has a store card or not.) **Hints:** In class, we provided solutions in R to a similar problem but for a different dataset.

(1) We will fit a logistic regression model with the UsesCoupon as the response and other variables Spends and HasCard as potential predictors using gradient descent. For this purpose, we will take the following steps:

(a) Let N be the number of data points we have. We first form the data matrix A whose first column is the vector of ones; second column is the Spends and third column is the HasCard vector. The matrix A should have size N 3. Let y be the UsesCoupon variable, which is a vector of zero and ones.

(b) The linear regression amounts to modelling the success probability as

$$\hat{y} = \sigma(Ax)$$

where $\sigma(z) = e^z/(1 + e^z)$ is the sigmoid function applied componentwise to the Ax vector so that

$$\hat{y}_i = \sigma\left(\sum_{j=1}^{3} A(i,j)x_j\right) = \sigma\left(x_1 + x_2[\mathbf{Spends}]_i + x_3[\mathbf{HasCard}]_i\right)$$

(0.1)

where $[\text{Spends}]_i$ and $[\text{HasCard}]_i$ are the i-th components of the vectors Spends and HasCard respectively. The aim is to choose $x = [x_1, x_2, x_3]^T$ (which are the coefficients of our model) so that the predicted probabilities $\hat{y}$ are close to the actual probabilities y as we discussed in class. For this purpose, the cost function to minimize in logistic regression is

$$f(x) = -\frac{1}{N}\left[\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right].$$

It follows that the j-th component of the gradient of the cost function $f(x)$ is

$$[\nabla f(x)]_j = \frac{\partial f}{\partial x_j} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)A(i,j), \quad i = 1, 2, 3$$

where $\hat{y}_i$ is given by (0.1), $A(i, j)$ is the entry of the A matrix at row i and column j. Implement the gradient descent method

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k)$$

to compute a minimum of f where you will be tuning the stepsize α_k to thedataset. Note that x^* is a column vector of size 3.

(2) Let $x^* = [x_1^*, x_2^*, x_3^*]^T$ be the minimum of f you computed from the previous question. The vector x^* is the coefficients you need for your predictive model. In
other words, you can predict the probability that a customer uses a coupon basedon the formula

$$\hat{y} = \sigma(x_1^* + x_2^*\text{Spends} + x_3^*\text{HasCard})$$

if you know the spending level of a customer or whether the customer has a card. For example, if the customer is spending 1000 dollars (then the variable Spends = 1) and he has a store card; your prediction will be

$$\hat{y} = \sigma(x_1^* + x_2^* 1 + x_3^*).$$

Make a table that contains the estimated (fitted) probabilities of your model versus the predictors HasCard and Spends. Your table should look like the table below where you will be filling out the missing values. Does having a store card increase the likelihood of the customers' using the coupons?

		Annual Spending						
		$1000	$2000	$3000	$4000	$5000	$6000	$7000
Credit Card	Yes							
	No							