ELE 364: Assignment #4

1. (10 pts) Consider the following multivariate linear regression model:

   Predict($\mathbf{d}$) = $\mathbf{w}[0]$ + $\mathbf{w}[1]\times\mathbf{d}[1]+\mathbf{w}[2]\times\mathbf{d}[2]+\mathbf{w}[3]\times\mathbf{d}[3]$

   and a historical dataset:

ID	$\mathbf{d}[1]$	$\mathbf{d}[2]$	$\mathbf{d}[3]$	Target
1	1	2	3	27
2	2	3	4	39
3	3	4	5	51
4	4	5	6	63

   (a) Assume we initialize the weights as follows: $\mathbf{w}[i] = i$ for i =0, 1, 2, 3. Predict the target values for all the instances in the dataset. Then, derive the value of the sum of squared error function L_2.

   (b) Assume a learning rate of 0.0001. Calculate the new $\mathbf{w}[i]$'s for i =0, 1, 2, 3 after one iteration.

2. (10 pts)

   A multivariate logistic regression model has been built to predict the propensity of shoppers to perform a repeat purchase of a free gift that they are given. The descriptive features are age of the customer, socio-economic band to which the customer belongs (a, b or c), average amount of money the customer spends on each visit to the shop, and the average number of visits the customer makes to the shop per week. The model is used by the marketing department to determine who should be given the free gift.

   When building multivariate logistic regression models, it is recommended that all continuous descriptive features be normalized to the range $[-1, 1]$. The table below shows a data quality report for the dataset used to train the model.

Feature	Count	% Miss.	Card.	Min.	1^{st} Qrt.	Mean	Median	3^{rd} Qrt.	Max.	Std. Dev.
Age	5,000	5	30	18	22	32.7	32	32	63	12.0
Shop value	5,000	0	3,000	5	12	101.2	100.1	174.9	230.7	71.1
Shop freq.	5,000	0	300	0.2	1	2.2	1.3	4.2	5.4	1.7

Feature	Count	% Miss.	Card.	Mode	Mode Count	Mode %	2^{nd} Mode	2^{nd} Mode Count	2^{nd} Mode %
Soc.-econ. band	5,000	6	3	a	2,600	52.0	b	1,315	26.3
Repeat purchase	5,000	0	2	no	2,820	56.4	yes	2,180	43.6

   Based on the above information, all continuous features were range-normalized and missing values replaced using mean for continuous features and mode for categorical features. Then, a multivariate logistic regression model was trained to obtain the weights shown below.

Feature	Weight
Intercept (**w[0]**)	0.668
Age	−0.580
Soc.-econ. band *a*	−0.160
Soc.-econ. band *b*	−0.198
Soc.-econ. band *c*	−0.232
Shop value	3.409
Shop freq.	2.050

ID	Age	Soc.-econ. band	Shop value	Shop freq.
1	?	*b*	204.62	1.33
2	62	?	110.50	0.85

Use this table to make predictions for each of the query instances shown in the table above (question marks refer to missing values).

3. (10 pts) End-of-chapter Problem 8 for ID = 1 query.

4. (10 pts) The use of the **kernel trick** is key to obtaining efficient implementations of the **support vector machine** approach to predictive modeling. The kernel trick is based on the fact that the result of a **kernel function** applied to a support vector and a query instance is equivalent to the result of calculating the dot product between the support vector and the query instance after a specific set of basis functions has been applied to both. In other words, $kernel(\mathbf{d},\mathbf{q}) = \phi(\mathbf{d})\cdot\phi(\mathbf{q})$.

   Consider the following $\phi(\langle d[1],d[2]\rangle)$ mapping:

   - $\phi_0(\langle d[1], d[2]\rangle) = d[1]^2$
   - $\phi_1(\langle d[1], d[2]\rangle) = d[2]^2$
   - $\phi_2(\langle d[1], d[2]\rangle) = \sqrt{2}\times d[1]\times d[2]$
   - $\phi_3(\langle d[1], d[2]\rangle) = \sqrt{2}\times d[1]$
   - $\phi_4(\langle d[1], d[2]\rangle) = \sqrt{2}\times d[2]$
   - $\phi_5(\langle d[1], d[2]\rangle) = 1$

   (a) Obtain the corresponding 6-dimensional vectors $\phi(v_1)$ for $v_1 = \langle 0.5, 0.8\rangle$, and $\phi(v_2)$ for $v_2 = \langle 1, 0.7\rangle$.

   (b) What is $\phi(v_1)\cdot\phi(v_2)$?

   (c) Suppose $kernel(\mathbf{d},\mathbf{q}) = (\mathbf{d}\cdot\mathbf{q}+1)^2$. What is $kernel(v_1, v_2)$? Does this match your previous calculation in (b)?

   (d) Given two support vectors:

   - $v_0 = \langle 0.9, 0.6\rangle$, class = −1
   - $v_1 = \langle 0.5, 0.8\rangle$, class = +1

   and their trained parameter set $w_0 = 0.21$, $\alpha_0 = 0.71$, and $\alpha_1 = 0.98$, predict the class of v_2 with this kernel function. Provide calculation details along with the results.

5. (20 pts) **Coding project**

   For this project, you will use error-based learning algorithms to train models that determine whether a credit card transaction is fraudulent.

   The dataset consists of numerical features, most of which have been derived through a preprocessing step (no additional background information is available on the original features because of confidentiality concerns). Another feature measures the amount of time elapsed between a transaction and the first transaction in the dataset, and the final feature measures the transaction amount.

   The target feature is a binary variable, where 1 indicates that the transaction is fraudulent.

   You will train a logistic regression classifier and support vector machines. You will also assess the impact of different hyperparameters, including different kernel functions and their parameters.

   See the Jupyter notebook for more details.