VE 445 2019 Fall Assignment 1

Question 1 (Linear Regression)

Some economist say that the impact of GDP in 'current year' will have effect on vehicle sales 'next year'. So whichever year GDP was less, the coming year sales was lower and when GDP increased the next year vehicle sales also increased.

•Let's have the equation as $y=\theta_0+\theta_1 x$, where

•y= number of vehicles sold in the year

•x= GDP of prior year

We need to find $\theta_0$and θ_1

•Here is the data between 2011 and 2016.

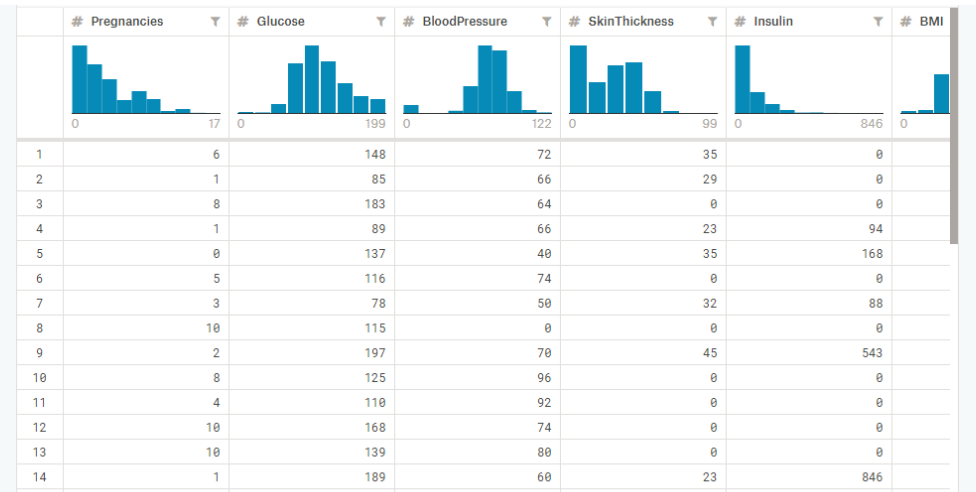
Year	GDP	Sales of vehicle
2011	6.2	
2012	6.5	26.3
2013	5.48	26.65
2014	6.54	25.03
2015	7.18	26.01
2016	7.93	27.9
2017		30.47
2018		

a) what is the normal equation?

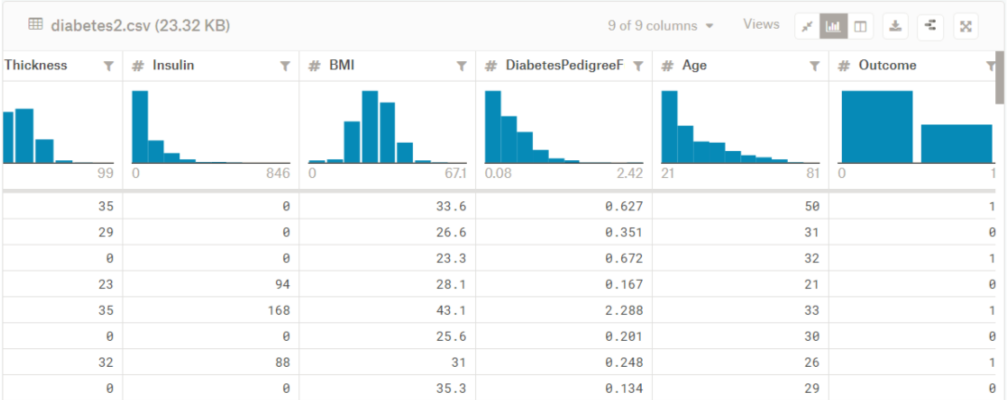
b) suppose the GDP increasement in 2017 is 7%, how many vehicles will be sold in 2018?

Question 2 (Logistic Regression)

National Institute of Diabetes and Digestive and Kidney Diseases wants to establish a model that automatically diagnose whether a Pima Indian has diabetes and provides a dataset as the material of the establishment of the model. The dataset contains information like Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, of female Pima Indians over 21. In the dataset, outcome labels as 1 means that the person has diabetes. Otherwise, she does not have, Implement Logistic Regression to establish this model, and use gradient decent to optimize the model. Change the learning rate gradient decent to plot accuracy-learning rate graph.



	# Pregnancies	# Glucose	# BloodPressure	# SkinThickness	# Insulin	# BMI
	0 – 17	0 – 199	0 – 122	0 – 99	0 – 846	0
1	6	148	72	35	0	
2	1	85	66	29	0	
3	8	183	64	0	0	
4	1	89	66	23	94	
5	0	137	40	35	168	
6	5	116	74	0	0	
7	3	78	50	32	88	
8	10	115	0	0	0	
9	2	197	70	45	543	
10	8	125	96	0	0	
11	4	110	92	0	0	
12	10	168	74	0	0	
13	10	139	80	0	0	
14	1	189	60	23	846	



Thickness	Insulin	BMI	DiabetesPedigreeF	Age	Outcome
35	0	33.6	0.627	50	1
29	0	26.6	0.351	31	0
0	0	23.3	0.672	32	1
23	94	28.1	0.167	21	0
35	168	43.1	2.288	33	1
0	0	25.6	0.201	30	0
32	88	31	0.248	26	1
0	0	35.3	0.134	29	0

Question 3 (SVM)

1. **Hyperplane:**

   A hyperplane in an N-dimension Euclidean space is defined as following:

$$\omega_1 x_1 + \omega_2 x_2 \cdots \omega_n x_n + b = 0$$

2. **SVM:**

   Given sample space $S = \{(\mathbf{x_1}, y_1)(\mathbf{x_2}, y_2), \cdots (\mathbf{x_m}, y_m), y_i = \{-1, 1\}\}$, we hope to find a hyperplane to divide the samples into two parts.

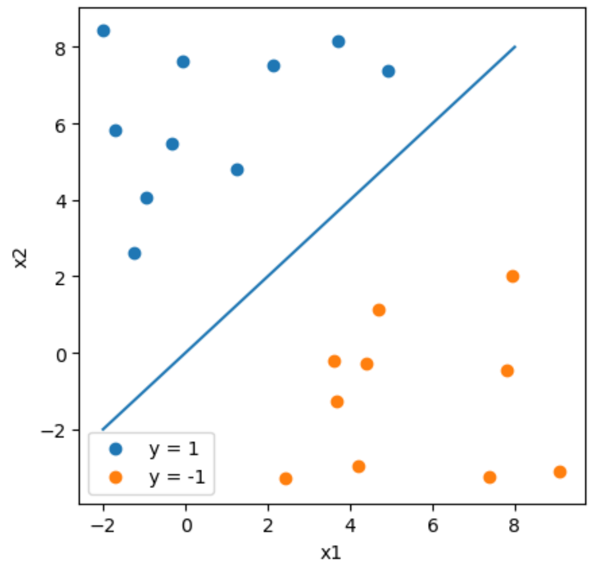


Figure 1. A division by hyperplane

Here the hyperplane can be defined as $\omega^{\mathbf{T}}\mathbf{x} + b = 0$, where $\omega = (\omega_1, \cdots, \omega_d)$, and b is the displacement. Assuming that the hyperplane (ω, b) can classify the sample correctly, then for each sample $(\mathbf{x_i}, y_i) \in S$, we have:

$$\begin{cases} \omega^T \mathbf{x} + b \geq 1, y_i = 1 \\ \omega^T \mathbf{x} + b \leq -1, y_i = -1 \end{cases}$$

The samples which make the equation achieved are called support vectors.

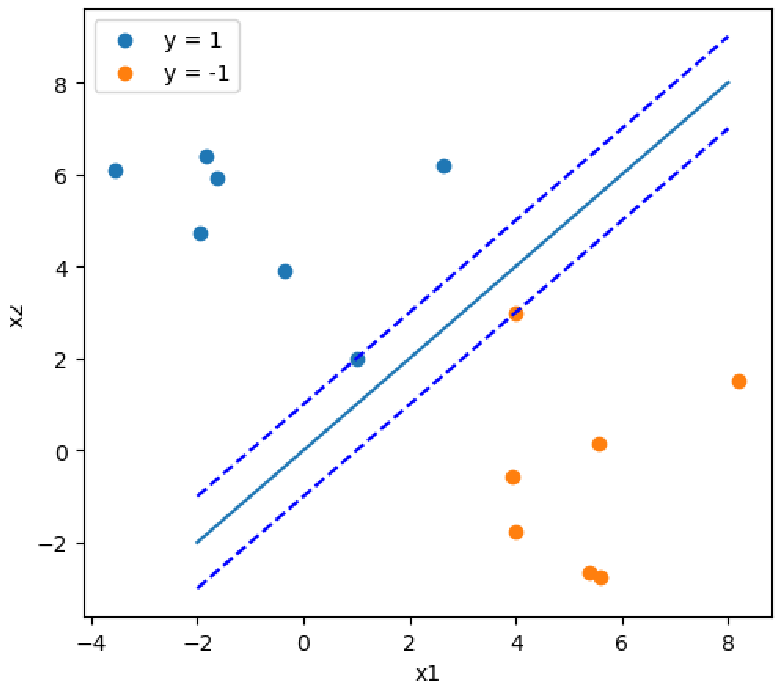


Figure 2. A margin to divide a sample space

The margin of the hyperspace is defined as the Euclidean distance between the hyperspace and two kinds of support vectors, which is given by:

$$\gamma = \frac{|\omega^T \mathbf{x} + b|_{y=1}}{||\omega||} + \frac{|\omega^T \mathbf{x} + b|_{y=-1}}{||\omega||} = \frac{2}{||\omega||}$$

Therefore the target function and constraint condition can be represented by:

$$\min_{\omega,b} \frac{1}{2}||\omega||^2 \qquad s.t. \quad y_i(\omega^T \mathbf{x_i} + b) \geq 1, \quad i = 1, 2, \cdots m.$$

By applying Lagrange multiplier method, its Lagrange function is given by:

$$\min_{\omega,b,\alpha} L(\omega, b, \alpha) = \frac{1}{2}||\omega||^2 + \sum_{i=1}^{m} \alpha_i (1 - y_i(\omega^T \mathbf{x} + b)), \quad \alpha_i \geq 0, \quad i = 1, \cdots, m$$

Set the derivation of the Lagrange function to be 0:

$$\omega = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x_i}, \quad \sum_{i=1}^{m} \alpha_i y_i = 0$$

The dual problem can be hence obtained:

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \mathbf{x_i^T} \mathbf{x_j}$$

$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \quad \alpha_i \geq 0, i = 1, 2, \cdots, m.$$

Using SMO algorithm, the function containing α can be optimized and the ω and b can be obtained by:

$$\omega = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x_i}, \quad y_s(\omega^T \mathbf{x_s} - b) \geq 1$$

where y_s and $\mathbf{x_s}$ is the label and feature of any support vector.

3. **Kernel SVM and Linear Non-separable Problem**:

   Sometimes in the original sample space it is impossible to find a hyperplane to divide the samples, which is known as linear non-separable.

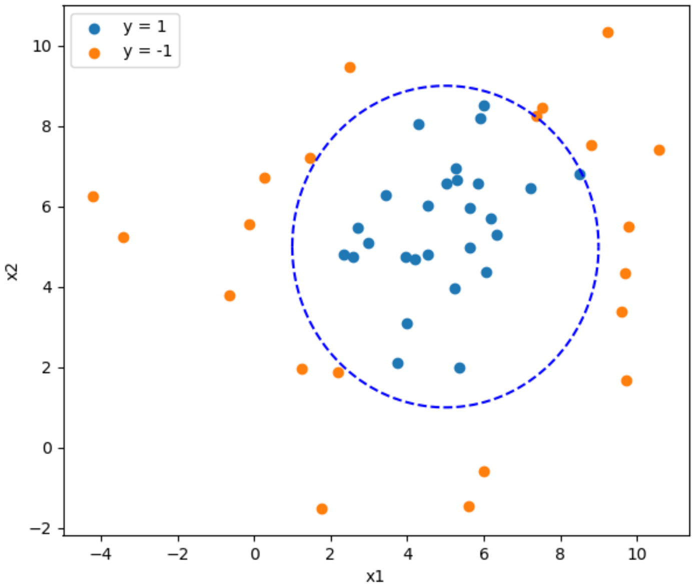


Figure 3. Linear Non-separable

For such case, a function $\phi(\mathbf{x})$ is expected to transfer the linear non-separable problem into a linear separable one, which maps the samples from lower dimension to higher dimensional Euclidean space, such that:

$$f(\mathbf{x}) = \omega^T\phi(\mathbf{x}) + b$$

And the target problem becomes:

$$\min_{\omega,b} \frac{1}{2}||\omega||^2 \qquad s.t. \quad y_i(\omega^T\phi(\mathbf{x_i}) + b) \geq 1, \quad i = 1, 2, \cdots m.$$

And the dual problem becomes:

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{m} \alpha_i\alpha_j y_i y_j \phi(\mathbf{x_i})^T\phi(\mathbf{x_j})$$

Here a kernel function is expected to simplify the inner product of higher dimension function $\phi(\mathbf{x})$, where $k(\mathbf{x_i}, \mathbf{x_j}) = \phi(\mathbf{x_i})^T\phi(\mathbf{x_j})$. Regardless of what $\phi(\mathbf{x})$ exactly is, we typically choose kernel function in the following:

Type	**Expression**	**parameter**
Linear kernel function	$k(\mathbf{x_i}, \mathbf{x_j}) = \mathbf{x_i}^T\mathbf{x_j}$	
Polynomial kernel function	$k(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i}^T\mathbf{x_j})^d$	$d \geq 1$
Gaussian kernel function	$k(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\vert\vert\mathbf{x_i} - \mathbf{x_j}\vert\vert/2\sigma^2)$	$\sigma > 0$
Laplace kernel function	$k(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\vert\vert\mathbf{x_i} - \mathbf{x_j}\vert\vert/\sigma)$	$\sigma > 0$

Table 1. Common kernel functions

And the original problem can be represented as:

$$f(\mathbf{x}) = \omega^T \phi(\mathbf{x}) + b = \sum_{i=1}^{m} \alpha_i y_i k(\mathbf{x}, \mathbf{x_i}) + b,$$

according to which the sample can be classified.

In the previous example, the following sample space can be generated by applying Gaussian function.

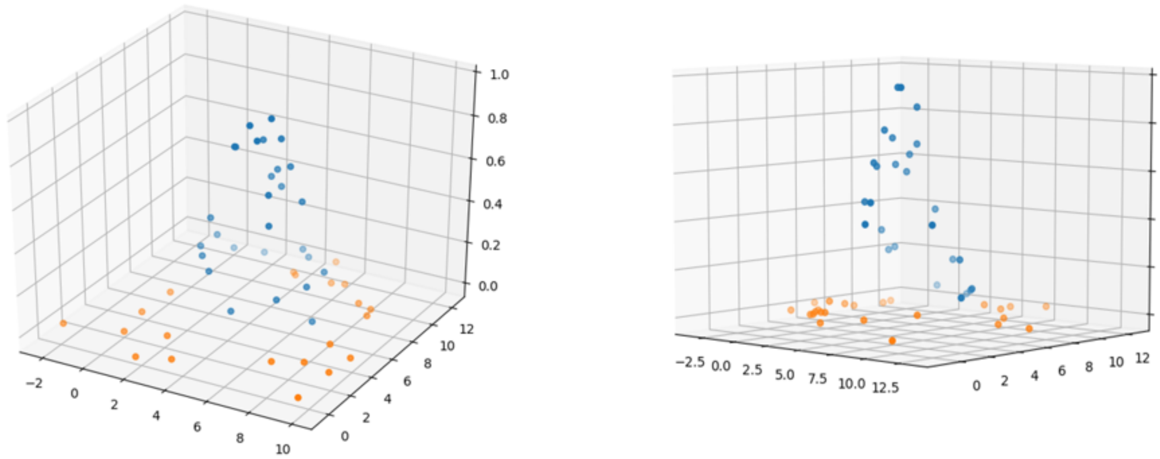


Figure 4. Higher dimensional space with Gaussian kernel function

4. **Soft Margin SVM**

   A soft margin SVM allows some samples that do not satisfy the hard margin constraint:

$$y_i(\omega^T \mathbf{x_i} + b) \geq 1, i = 1, 2, \cdots, m.$$

   In order to maximize the margin and minimize the unsatisfactory sample number, the optimization function can be represented as:

$$\min_{\omega,b} \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m} l(y_i(\omega^T \mathbf{x_i} + b) - 1) := \min_{\omega,b} \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m} \xi_i$$

   where $l(x)$ is a loss function and C is a constant. In this lab, we use the loss function:

$$l(x) = \max(0, 1 - x)$$

   Its Lagrange function is given by:

$$L(\omega, b, \alpha, \xi, \mu) = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i(1 - \xi_i - y_i(\omega^T \mathbf{x_i} + b)) - \sum_{i=1}^{m} \mu_i \xi_i$$

   Set the derivation of the Lagrange function to be 0:

$$\omega = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x_i}, \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \quad C = \alpha_i + \mu_i$$

   The dual problem is given by:

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j \mathbf{x_i^T x_j}$$

$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, i = 1, 2, \cdots, m.$$

It can be seen that the only difference between soft margin SVM and hard margin SVM is just the constraint of α . Similarly, this dual problem can be easily solved by SMO algorithm and any other optimization algorithm.

Specifics

1. Complete Question 1 a) and b).
2. Complete Question 2 by plotting the graph and attach your code. The dataset is Assign1_Q2.csv.
3. In Question 3, realize the four types of SVM, and strictly follow the given template Assign1_Q3.py.
4. Optimize the models by SMO, the details about SMO is attached.
5. Train the models based on Iris dataset introduced in class using SVM.
6. [20%] Write report about your training procedure and compare your result with the result by using library function provided by `sklearn` module or other python modules.

Submission

You should submit a .zip file to Canvas by Oct.15 11:59 pm, which is expected to contain the answers of Question 1 and 2, a file named SVM.py, and a report.