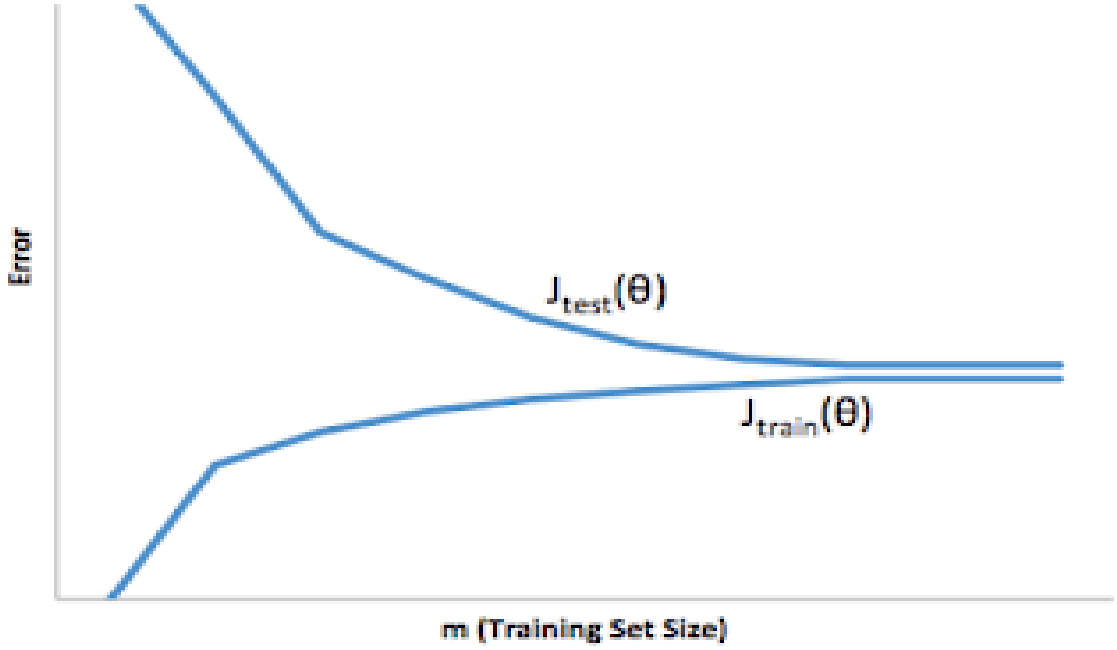
ML Module Bank-1

1. a. The plot shown in the figure has training and testing error of an algorithm. Is the algorithm suffering from high bias, high variance, or neither? Justify it with proper explanation?

b. Take any learning algorithm and train the model, and find that it has unacceptably high error on the test set. Plot the learning curve, and write observations?



2. A dataset has three class labels, namely Apple, Orange and Mango. Thefollowing is a possible confusion matrix for these classes.

		True class		
		APPLE	ORANGE	MANGO
Predicted Class	APPLE	8	6	7
	ORANGE	2	6	2
	MANGO	4	2	3

For reference:

Accuracy = (true positives + true negatives) / (total examples)

Precision = (true positives) / (true positives + false positives)

Recall = (true positives) / (true positives + false negatives)

F1 score = (2 * precision * recall) / (precision + recall)

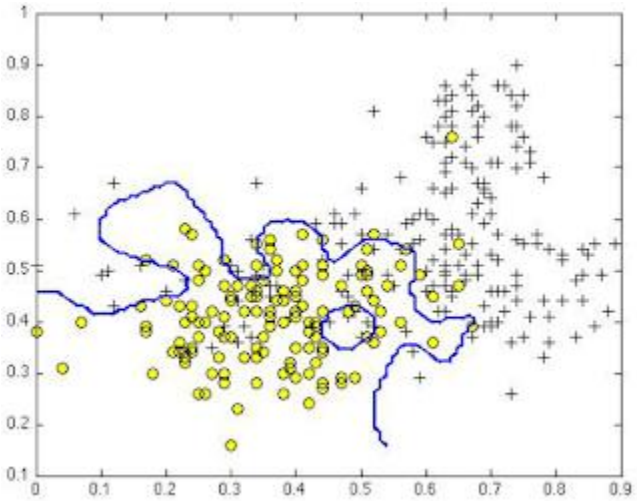
Find the class wise performance evaluation measures and analyze which class accuracy is more out three classes. Also, measure the algorithm for classifier's F1 score (as a value form 0 to 1)

3. Many substances that can burn (such as gasoline and alcohol) have a chemical structure based on carbon atoms; for this reason, they are called hydrocarbons. A chemist wants to understand how the number of carbon atoms in a molecule affects how much energy is released when that molecule combusts (meaning that it is burned). The chemist obtains the dataset below. In the column on the right, "kJ/mol" is the unit measuring the amount of energy released.

Name of molecule	Number of hydrocarbons in molecule (x)	Heat release when burned (kJ/mol) (y)
methane	1	-890
ethene	2	-1411
ethane	2	-1560
propane	3	-2220
cyclopropane	3	-2091
butane	4	-2878
pentane	5	-3537
benzene	6	-3268
cycloexane	6	-3920
hexane	6	-4163
octane	8	-5471
napthalene	10	-5157

You would like to use linear regression ($h_\theta(x) = \theta_0 + \theta_1 x$) to estimate the amount of energy released (y) as a function of the number of carbon atoms (x). What are values do you think to obtain for θ_0 and θ_1 ? Also, Plot the given points and the regression line in the same rectangular system of axes.

4. Suppose you have trained an SVM classifier with a Gaussian kernel, and it learned the following decision boundary on the training set:



When you measure the SVM's performance on a cross validation set, it does poorly. Should you try with increasing or decreasing C and also increasing or decreasing σ^2 ?

5. Suppose m=4 students have taken some classes, and the class had a midterm exam and a final exam. You have collected a dataset of their scores on the two exams, which is as follows:

midterm exam	(midterm exam)^2	final exam
89	7921	96
72	5184	74
94	8836	87
69	4761	78

You'd like to use polynomial regression to predict a student's final exam score from their midterm exam score. Concretely, suppose you want to fit a model of the form $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, where x1 is the midterm score and x_2 is (midterm score)^2. Further, you plan to use both feature scaling (dividing by the "max-min", or range, of a feature) and mean normalization. What is the normalized feature $x_2^{(4)}$? (Hint: midterm = 69, final = 78 is training example 4.)

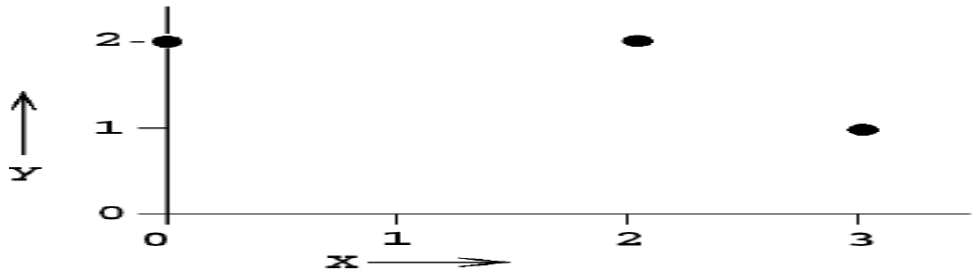
6. Consider the problem of predicting how well a student does in her second year of college/university, given how well she did in her first year. Specifically, let x be equal to the number of "A" grades (including A-. A and A+ grades) that a student receives in their first year of college (freshmen year). We would like to predict the value of y, which we define as the number of "A" grades they get in their second year (sophomore year). Here each row is one training example. Recall that in linear regression, our hypothesis is $h\theta(x) = \theta_0 + \theta_1 x$ to denote the number of training examples.

X	Y
3	3
2	3
0	1
3	2

For the training set given above,

a. Find the least square regression line for the given data points.
b. Plot the given points and the regression line in the same rectangular system of axes.
c. Use the least square regression line as a model to estimate **y** when x=12.

7. Suppose you have this data set with one real-valued input and one real-valued output:



x	y
0	2
2	2
3	1

a. What is the mean squared leave one out cross validation error of using linear regression? (i.e. the mode is y = _0 + _1x + noise)
b. Suppose you use a trivial algorithm of predicting a constant y = c. What is the mean squared leave one out error in this case? (Assume c is learned from the non-left-out data points.)

8. Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines. You model uses $p(x) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$

You have two features x_1 = vibration intensity, and x_2 = heat generated. Both x_1 & x_2 take on values between 0 and 1 (and are strictly greater than 0), and for most "normal" engines you expect $x_1 \approx x_2$. One of the suspected anomalies is that a flawed engine may vibrate very intensely even without generating much heat (large x_1, small x_2), even though the particular values x_1 & x_2 may not fall outside their typical ranges of values. What additional feature x_3 should you create to capture these types of anomalies?

9. Suppose you are given the following set of data with three Boolean input variables a; b; and c, and a single Boolean output variable K.

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

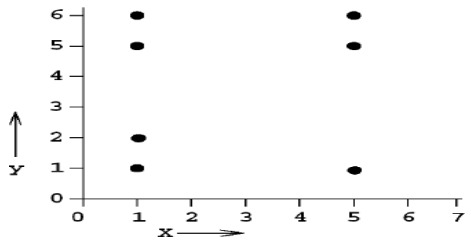
For parts (a) and (b), assume we are using a naive Bayes classifier to predict the valueof K from the values of the other variables.

a. According to the naive Bayes classifier, Comprehend P(K = 1|a = 1 ^ b = 1 ^ c = 0)?
b. According to the naive Bayes classifier, Comprehend P(K = 0|a = 1 ^ b = 1)?

Now, suppose we are using a joint Bayes classifier to predict the value of K from the values of the other variables.

c. According to the joint Bayes classifier, Comprehend P(K = 1|a = 1 ^ b = 1 ^ c = 0)?
d. According to the joint Bayes classifier, Comprehend P(K = 0|a = 1 ^ b = 1)?

10. The following picture shows a dataset with one real-valued input x and one real-valued output y. There are seven training points.



Suppose you are training using kernel regression using some unspecified kernel function. The only thing you know about the kernel function is that it is a monotonically decreasing function of distance that decays to zero at a distance of 3 units (and is strictly greater than zero at a distance of less than 3 units).

a. Compute and analyze predicted value of y when x = 1?
b. Compute and analyze predicted value of y when x = 3?
c. Compute and analyze predicted value of y when x = 4?
d. Compute and analyze predicted value of y when x = 7?