Name and Student ID:

Machine Learning BLG527E, Nov 6, 2014, 120mins, Midterm Exam (PART A)
Signature:
Duration: 120 minutes.
Closed books and notes. Write your answers neatly in the space provided for them. Write your name on each sheet. Good Luck!

Q1	**Q2**	**Q3**	**Q4**	**Q5**	Q6	Q7	Q8	Q9	TOTAL
10	**6**	**14**	**15**	**10**	10	10	10	15	100

QUESTIONS

Q1) [10pts]
In the table below, x_1, x_2, x_3 and $x_i \in \{0,1\}$, $i = 1,2,3$ x_i represent the *i* feature vector and $y \in \{+,-\}$ represents the class label.

Id	x_1	x_2	x_3	**y**
1	1	0	0	-
2	0	1	0	-
3	0	0	1	-
4	0	0	0	+
5	1	1	1	+

1a) Construct the Naïve Bayes classifier for the given training dataset.

1b) Classify the $(x_1 = 1, x_2 = 1, x_3 = 0)$ data sample.

Q2) [6pts]
Suppose you are given a financial regression dataset generated from a polynomial of degree of 4. Indicate whether you think the bias and variance of the following models would be relatively high (H) or low (L) considering the true model.

	Bias	Variance
Linear Regression		
Polynomial regression with degree of 4		
Polynomial regression with degree of 9		

Name and Student ID:

Q3) [14pts]

3a) Generate a decision tree using Gini index (2p(1-p)) as impurity measure.

Weekend (x_1)	**Rain (x_2)**	**Daytime (x_3)**	**Take Taxi C**
Yes	No	Morning	+
Yes	Yes	Morning	+
Yes	Yes	Morning	+
No	Yes	Evening	+
No	Yes	Evening	+
No	No	Noon	-
Yes	No	Noon	-
Yes	Yes	Noon	-
No	No	Evening	-
No	No	Evening	-

3b) Are there any irrelevant features?

Name and Student ID:

Q4)[15pts]

The probability of a single observation x with rate parameter θ follows the following Poisson distribution:

$$\boldsymbol{P}(\boldsymbol{x}|\theta) = \frac{\theta^x e^{-\theta}}{x!}, for\ x = 0,1,2 \ldots n$$

You are given the data points x_1 x_2,… x_n that are drawn independently from Poisson distribution with parameter θ.

Write down the log-likelihood of the data:

Find the maximum likelihood estimate of the parameter θ:

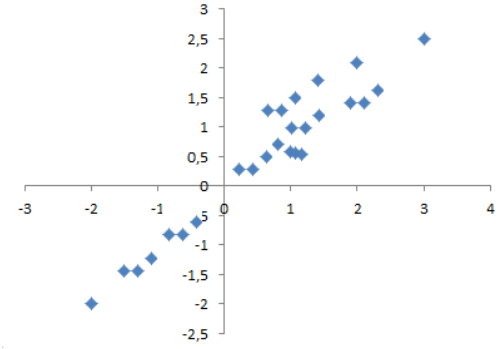
Q5)[10pts]

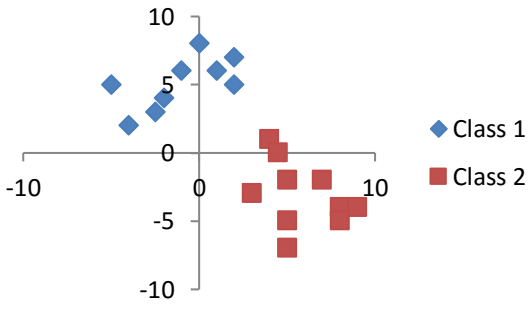
Describe briefly the following dimensionality reduction methods:

Principal Component Analysis (PCA):

Linear Discriminant Analysis (LDA):

Plot the directions of the first PCA and LDA components of the following figures.





Name and Student ID:

If you need more space, use this page only for PART A (Q1-Q5).

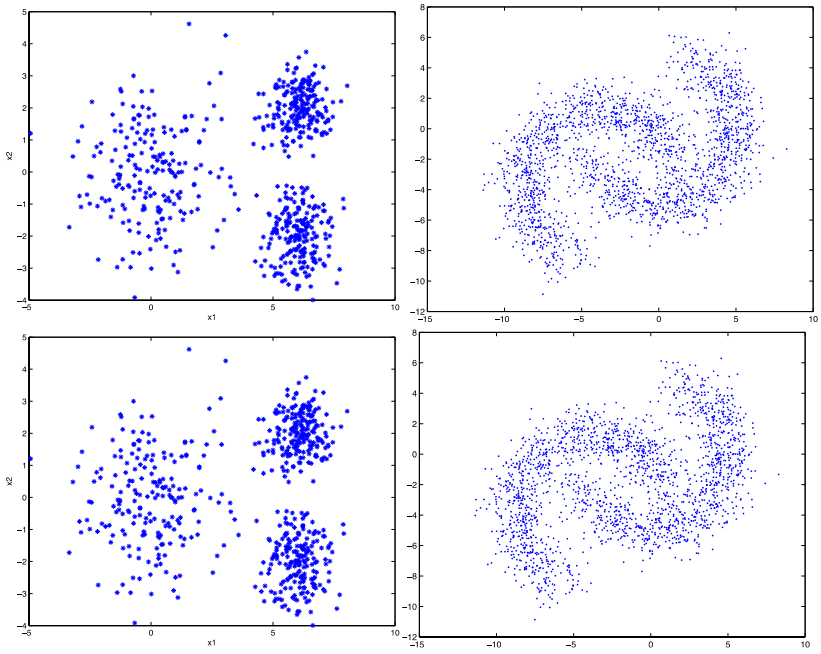
Name and Student ID:

Machine Learning BLG527E, Nov 6, 2014, 120mins, Midterm Exam (PART B)
Signature:
Duration: 120 minutes.
Closed books and notes. Write your answers neatly in the space provided for them. Write your name on each sheet. Good Luck!

Q1	*Q2*	*Q3*	*Q4*	*Q5*	**Q6**	**Q7**	**Q8**	**Q9**	TOTAL
10	*6*	*14*	*15*	*10*	**10**	**10**	**10**	**15**	100

Q6)[10pts]
Briefly explain the K-Means algorithm:

What are two possible clusterings of the following datasets using K-Means with K=4. Only draw the final clusters.



Considering these datasets, does the K-Means algorithm have any drawbacks? Can you suggest alternative clustering methods?

Name and Student ID:

Q7)[10pts]
What is the VC dimension of a line in 2 dimensional space? Explain your answer.

Q8)[10pts]
Assume that g is a quadratic model and for input x, z=[x_1^2 x_2^2 …x_d^2]

which outputs the following:

$$g(\mathrm{x}, \mathrm{v}, \mathrm{w}, w_0) = \mathrm{v}^{\mathrm{T}}\mathrm{z} + \mathrm{w}^{\mathrm{T}}\mathrm{x} + w_0$$

You need to make the parameters v of g take smaller values and hope that it will result in better generalization. Given a dataset $X = \{\mathrm{x}^t, r^t\}_{t=1}^N$, how would you obtain the solution for v,w and w_0 in this situation.

Hint: Modify the sum of squares error function to incorporate the need of smaller v values, and derive the solution for v,w and w_0 analytically.

Name and Student ID:

Q9)[15pts]
Suppose you became the head of the department in the future and students have complained that amount of cheese (kaşar) in the sandwiches of the canteen has significantly reduced. You know that according to the regulations the cheese should be 50grams. After the complaints you go to the canteen and buy 10 sandwiches and measure the amount of cheese in them as follows:

51	51	51	51	51	47	47	47	47	47

Can you decide with 95% confidence that amount of cheese is significantly reduced? Justify your decision.

Hint:
If variance is known, with m average of x_i , we accept that x is less than p_0 with 100(1-α) confidence if

$\frac{\sqrt{N}(m - m_0)}{s} \sim z_a$ is less than z_α

If variance is not known, with m and s^2 average and var of x_i , we accept that x is less than p_0 with 100(1-α) confidence if

$\frac{\sqrt{N}(m - p_0)}{S} \sim t_{N-1}$ is less than $t_{\alpha,N-1}$

$t_{0.05,9} = 1.83$, $t_{0.025,9} = 2.26$, $z_{0.05} = 1.64$, $z_{0.025} = 1.96$, choose wisely.

In order to make your computations easier, when computing the standard deviation, divide by N (not by N-1). Take $\sqrt{10}$ =3.16.

Name and Student ID:

If you need more space, use this page only for PART B (Q6-Q9).