

**ITU, Computer Engineering Dept.  
BLG527E, Machine Learning HW1**

Due: November 17, 2024, 23:00 through Ninova.

**NO LATE SUBMISSION WILL BE ACCEPTED. DO NOT SUBMIT THROUGH E-MAIL.**

Instructors: Ayşe Tosun Kühn ([tosunay@itu.edu.tr](mailto:tosunay@itu.edu.tr)) Yusuf Yaslan ([yvaslan@itu.edu.tr](mailto:yvaslan@itu.edu.tr))

**Grading: You must complete the table below according to what you expect to get out of each question. At the beginning of your report you should give the following table. You can use the given report template.**

		Q1	Q2	Q3	Q4	Q5	Total
<b>Grade</b>	<b>Max</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>0 pts</b>
	<b>Expected</b>						

**Policy:**

Please do your homeworks on your own. You are encouraged to discuss the questions with your class mates, but the code and the hw you submitted must be your own work. Cheating is highly discouraged for it could mean a zero or negative grade from the homework.

If a question is not clear, please let us know (via email or in class). Unless we indicate otherwise, do not use libraries for machine learning methods. When in doubt, email me.

There will be 3 homeworks this term. Each hw is worth 10 points and each question will be evaluated on a 0/1 basis.

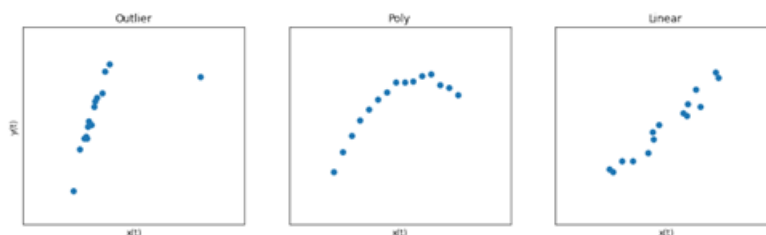
Q1) Make sure that you read Appendix A of the textbook and the resources on matrices, linear algebra and probability and statistics on Ninova.

Given the table of joint probabilities between three discrete random variables X, Y, and Z

P(X,Y,Z)	X	0			1		
	Y	-1	0	1	-1	0	1
Z	1	0.06	0.06	0.08	0.06	0.02	0.02
	2	0.09	0.09	0.12	0.24	0.08	0.08

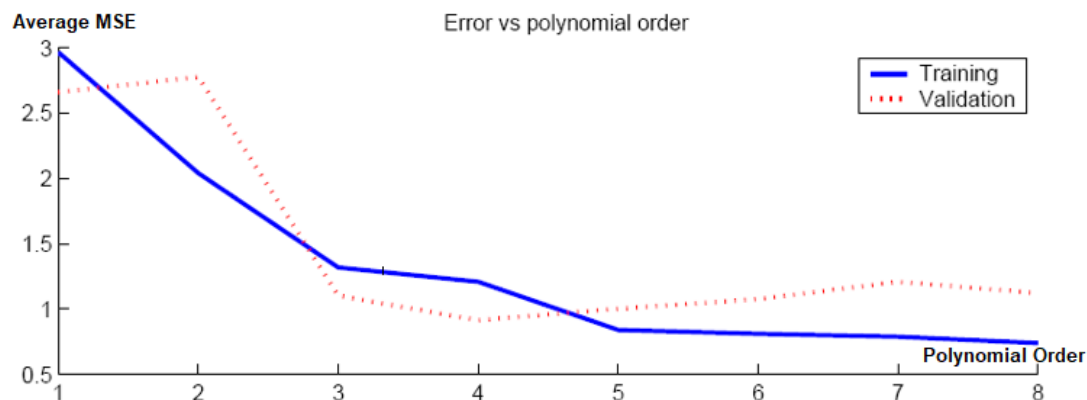
- Evaluate the joint probability mass function  $P(Z/X=0, Y=0)$ .
- Are random variables Z and Y independent? Show your solution. (Hint: Marginalize X)

Q2) Do this exercise in python. Solve a) and b) using closed form solution



- Given the following data points, compute the least squares regression line and plot it with red color on top of the data points for all three cases.
- For each dataset apply 5 fold cross validation (Shuffle the data before splitting into the folds) for different orders of polynomial fits (From 1<sup>st</sup> order to 8<sup>th</sup> order).  
For each order of polynomial you will do the following:  
Obtain training MSE and validation MSE for each fold. Then Compute average training MSE and average Validation MSE values.

For each dataset give the average error vs polynomial order plot. An example plot is depicted below.



Make sure that for every order of polynomials you use the same folds.  
Discuss what may happen if the folds are different for different models.  
Discuss what may happen if the folds are obtained without shuffling.

Q3) A pharmaceutical company prepared a test amid the Coronavirus outbreak. The test is able to predict positive when applied to subjects with the virus with 90% accuracy while it predicts negative for healthy subjects again with 90% accuracy. We also know that only 2% of the population has the virus. What is the probability of actually being sick if the test predicts positive? Would you trust this test? A pharmaceutical company prepared a test amid the Coronavirus outbreak. The test is able to predict positive when applied to subjects with the virus with 90% accuracy while it predicts negative for healthy subjects again with 90% accuracy. We also know that only 2% of the population has the

virus. What is the probability of actually being sick if the test predicts positive? Would you trust this test?

Q4) Let  $X$  and  $Y$  be two random variables, denoting age and weight, respectively. Consider a random sample of size  $n = 20$  from these two variable

$X = (69, 74, 68, 70, 72, 67, 66, 70, 76, 68, 72, 79, 74, 67, 66, 71, 74, 75, 75, 76)$

$Y = (153, 175, 155, 135, 172, 150, 115, 137, 200, 130, 140, 265, 185, 112, 140, 150, 165, 185, 210, 220)$

- a) Find the mean, median, and mode for  $X$ .
- b) What is the variance for  $Y$ ?
- c) Plot the normal distribution for  $X$ .
- d) What is the probability of observing an age of 80 or higher?
- e) Find the 2-dimensional mean  $\hat{\mu}$  and the covariance matrix  $\hat{\Sigma}$  for these two
- f) variables.
- g) What is the correlation between age and weight?
- h) Draw a scatterplot to show the relationship between age and weight.

Q5) a) A continuous random variable  $x$  has the Laplace distribution

$$p(x) = \frac{1}{2} \exp(-|x|)$$

- Plot  $p(x)$
- What is the probability that  $x > 2$ ?

b) A discrete random variable has the binomial distribution

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

What is the  $E[x] = ?$  and  $E[x^2] = ?$  Show their derivations step by step.