



ITU Computer and Informatics Faculty
BLG 454E Learning From Data, Spring 2018
Homework #1
Due March 07, 2018 10pm

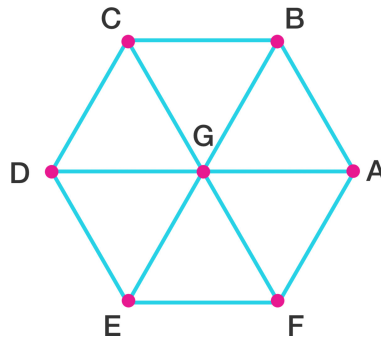
1. (10 pts.) In general, the probability that it rains on Saturday is 25%.

Weekend rain has the following relationships:

- If it rains on Saturday, the probability that it rains on Sunday is 50%.
- If it does not rain on Saturday, the probability that it rains on Sunday is 25%.

Given that it rained on Sunday, what is the probability that it rained on Saturday?

2. (20 pts.) A bug stands on a random point of the lattice below. Each point is equally likely to be the starting point.



Every minute, the bug selects an adjacent point at random and moves to it. Each adjacent point is equally likely to be chosen. For example, if the bug is on point B, then each probability to move to the points A, C, or G is $\frac{1}{3}$.

What is the probability that the bug reaches point A in 2 moves or less? *Each point is equally likely to be the bug's starting point. Also, assume starting at A will "reach" the point in 0 moves.*

3. (40 pts.) The idea for the **maximum likelihood estimate (MLE)** is to find the value of the parameter(s) for which the data has the highest probability. You are going to do this with the densities. Suppose the 1-dimension data points x_1, x_2, \dots, x_n given in "data.txt" file are drawn from a normal(gauss) $N(\mu, \sigma^2)$ distribution, where μ and σ are unknown.
- (20 pts.) Formulate the likelihood function and derive the equation to find the maximum likelihood estimate for the pair (μ, σ^2) .
 - (20 pts.) Implement (write the code) MLE in **Matlab** or **Python** language and provide your plot that is similar to Figure 1. **You are not allowed use any built-in functions except histogram functions to provide you a quick of the distribution of the data.**

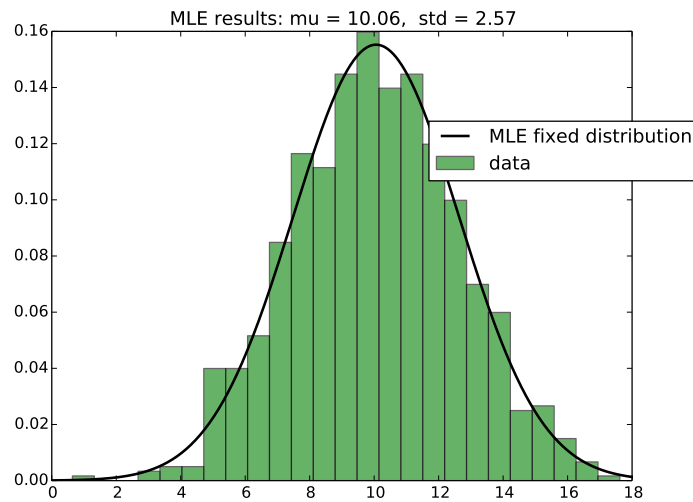


Figure 1: Data and fixed gaussian distribution with MLE

4. (30 pts.) In the Table 1 below, \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 and $\mathbf{x}_i \in \{0, 1\}$, $i = 1, 2, 3$. \mathbf{x}_i represent the i feature vector and $\mathbf{y} \in \{+, -\}$ represents the class label.
- (15 pts.) Construct the Naive Bayes classifier for the given training dataset in Table 1.
Hint: Estimate the class conditional prob. for each feature vector \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3
 - (5 pts.) Predict the class label for $(x_1 = 1, x_2 = 1, x_3 = 1)$ data using trained Naive Bayes approach in part (a)
 - (10 pts.) Calculate the probabilities of $P(x_1 = 1)$, $P(x_2 = 1)$, and $P(x_1 = 1, x_2 = 1)$. Decide whether x_1 and x_2 are independent or not.

Table 1: Training set for question 4

| Instance | x_1 | x_2 | x_3 | \mathbf{y} |
|----------|-------|-------|-------|--------------|
| 1 | 0 | 0 | 1 | - |
| 2 | 1 | 0 | 1 | + |
| 3 | 0 | 1 | 0 | - |
| 4 | 1 | 0 | 0 | - |
| 5 | 1 | 0 | 1 | + |
| 6 | 0 | 0 | 1 | + |
| 7 | 1 | 1 | 0 | - |
| 8 | 0 | 0 | 0 | - |
| 9 | 0 | 1 | 0 | + |
| 10 | 1 | 1 | 1 | + |

Submission Policy

- Prepare the report and code. Only electronic submissions through Ninova will be accepted no later than March, 07 at 10pm.
- You may discuss the problems at an abstract level with your classmates, but you should not **share or copy code** from your classmates or from the Internet. You should submit your **own, individual** homework.
- Academic dishonesty, including cheating, plagiarism, and direct copying, is unacceptable.
- Note that **your codes and reports will be checked with the plagiarism tools!**

- If a question is not clear, please let the teaching assistants know by email (kivrakh@itu.edu.tr or cebeci16@itu.edu.tr).

Bonus marks (10pts)

- Clarity and nicely described report
- Using Latex template for the report that is given to you

Deductions (-10pts)

- Spelling errors.
- Messiness
- Lack of content.
- Irrelevant / mistaken content.