OSU CSE 3521
Homework #3: Problem Set

Release Date: October 26th, 2021

Submission Instructions

Due Date: November 12th (23:59 ET), 2021

Submission: Please submit your solutions in a single PDF file named HW3_name_number.pdf (e.g., HW3_chao_209.pdf) to Carmen. You may write your solutions on paper and scan it, or directly type your solutions and save them as a PDF file. *Submission in any other format will not be graded.*

We highly recommend that you write down the derivation of your answers, and highlight your answers clearly!

Collaboration: You may discuss with your classmates at a very high level. However, you need to write your own solutions and submit them separately by yourself. Also in your written report, you need to list with whom you have discussed for each problem (please do so in the first page). Please consult the syllabus for what is and is not an acceptable collaboration.

Calculation: Please perform rounding to your results after the second decimal number, *unless stated otherwise*. For example, 1.245 becomes 1.25 and -1.228 becomes -1.23.

1 Naive Bayes and MLE [11 points]

	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{x}_4$	$\boldsymbol{x}_5$	$\boldsymbol{x}_6$	$\boldsymbol{x}_7$	$\boldsymbol{x}_8$	$\boldsymbol{x}_9$	$\boldsymbol{x}_{10}$
$\boldsymbol{x}_i[1]$	1	1	1	1	0	0	0	0	0	1
$\boldsymbol{x}_i[2]$	1.0	2.0	2.5	3.5	4.0	5.0	0.5	1.0	2.0	2.5

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
0	0	0	0	0	0	1	1	1	1

Figure 1: A two-dimensional labeled dataset with 10 data instances.

Figure 1 gives a labeled dataset $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$ of 10 data instances. Each $\boldsymbol{x}_i$ is two-dimensional: the first dimension $\boldsymbol{x}_i[1]$ is binary (i.e., $\{0, 1\}$); the second dimension $\boldsymbol{x}_i[2]$ is a real number. The label y_i is binary (i.e., either class 0 or class 1).

Denote by X the two-dimensional random variable of data instances and Y the binary random variable of class labels, you are to construct the Naive Bayes classifier to predict the class label $\hat{y}$ of a future data instance $X = \boldsymbol{x}$

$$\begin{aligned}\hat{y} &= \arg\max_{c\in\{0,1\}} p(Y = c|X = \boldsymbol{x}) \\ &= \arg\max_{c\in\{0,1\}} p(Y = c) \times \prod_{d=1}^{2} p(X[d] = x[d]|Y = c).\end{aligned}$$

You will begin by estimating the parameters of

- $p(Y)$
- $p(X[d]|Y = c) \quad \forall c \in \{0, 1\}, d \in \{1, 2\}$

from the labeled dataset, using **Maximum Likelihood Estimation (MLE)**. Note that, $p(Y)$ is a Bernoulli distribution; $p(X[1]|Y = c)$ is a Bernoulli distribution of $X[1]$; $p(X[2]|Y = c)$ is a Gaussian distribution of $X[2]$.

1.1 Prior distributions [1 points]

What is $p(Y = 1)$? The answer is a real number.

1.2 Conditional distributions [2 points]

What is $p(X[1] = 1|Y = 0)$?

What is $p(X[1] = 1|Y = 1)$?

Each answer is a real number.

1.3 Conditional distributions [4 points]

What is $p(X[2] = 3.0|Y = 0)$?

What is $p(X[2] = 3.0|Y = 1)$?

Each answer is a real number. Please first write down the parameters (i.e., mean and variance) of $p(X[2]|Y = 0)$ and $p(X[2]|Y = 1)$.

1.4 Naive Bayes [4 points]

Given $\boldsymbol{x} = [0, 3.0]^\top$ (i.e., $x[1] = 0$ and $x[2] = 3.0$), what is the prediction $\hat{y}$?

Given $\boldsymbol{x} = [1, 0.5]^\top$ (i.e., $x[1] = 1$ and $x[2] = 0.5$), what is the prediction $\hat{y}$?

Each answer is either 0 or 1.

Please do write down your derivation/calculation.

2 Probabilistic Graphical Models (PGMs) [14 points]

2.1 Joint distributions [3 points]

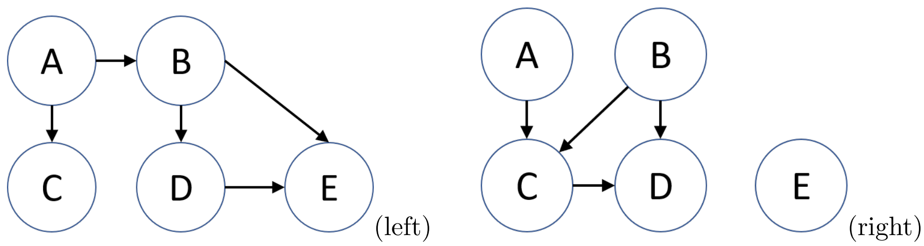


Figure 2: Two probabilistic graphical models (PGMs).

Figure 2 shows two PGMs of five random variables A, B, C, D, and E. The joint distribution $p(A, B, C, D, E)$ according to the PGM can be decomposed as follows:

$$\prod_{X \in \{A,B,C,D,E\}} p(X|\text{parents}(X))$$

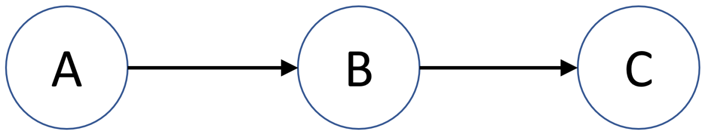
(a) Please decompose $p(A, B, C, D, E)$ for the left PGM in the above-mentioned form. The expected answer is an expression that solely comprises of probabilities over one or more of the random variables A, B ,C, D, E. For example, $p(A, B, C, D, E) = p(C)p(A|C)p(B|A, C)p(D|A, B, C)p(E|A)$.

(b) Please decompose $p(A, B, C, D, E)$ for the right PGM in the above-mentioned form. The expected answer is an expression that solely comprises of probabilities over one or more of the random variables A, B ,C, D, E.

2.2 PGMs from joint distribution decomposition [2 points]

Please draw the PGM of five random variables A, B, C, D, and E, given that the joint distribution $p(A, B, C, D, E) = p(A)p(B)p(C)p(D|A, B)P(E|D, C)$.

2.3 Inference [9 points]



$A =$	$p(A)$
a	0.8
$\neg a$	0.2

$B =$	$A =$	$p(B\|A)$
b	a	0.6
$\neg b$	a	0.4
b	$\neg a$	0.3
$\neg b$	$\neg a$	0.7

$C =$	$B =$	$p(C\|B)$
c	b	0.5
$\neg c$	b	0.5
c	$\neg b$	0.1
$\neg c$	$\neg b$	0.9

Figure 3: A probabilistic graphical model (PGM) of three random variables and the corresponding probability terms.

Figure 3 shows a PGM of three binary random variables A, B, and C, together with the corresponding probability terms. Each binary random variable has two outcomes. Please compute the following probabilities (answer must be a floating point decimal) and answer the following questions with a response of "Yes" or "No".

1. $p(B = b)$ [1 point]
2. $p(C = \neg c)$ [1 point]
3. $p(A = a, C = \neg c)$ [1 point]
4. Is $p(A = a, C = \neg c)$ equal to $p(A = a)p(C = \neg c)$? [1 point]
5. $p(A = a|B = b)$ [1 point]
6. $p(A = a, C = \neg c|B = b)$ [1 point]
7. Is $p(A = a, C = \neg c|B = b)$ equal to $p(A = a|B = b)p(C = \neg c|B = b)$? [1 point]
8. $p(A = a|C = \neg c)$ [2 points]

Hint:

- $P(C = \neg c) = P(C = \neg c, B = b) + P(C = \neg c, B = \neg b)$
  $= P(C = \neg c|B = b)p(B = b) + P(C = \neg c|B = \neg b)p(B = \neg b)$
- $P(A = a, C = \neg c) = P(A = a, C = \neg c, B = b) + P(A = a, C = \neg c, B = \neg b)$

Please do write down your derivation/calculation.

3 Conditional Independence [5 points]

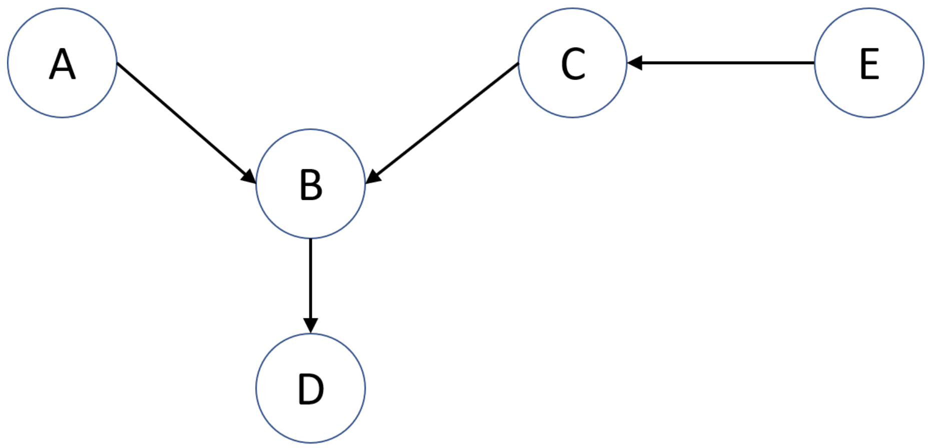


Figure 4: A probabilistic graphical models (PGM) of five random variables.

Figure 4 shows a PGM of five random variables A, B, C, D, and E. Please answer the following questions.

1. Is $A \perp E$ (i.e., $p(A, E) = p(A)P(E)$)?
2. Is $A \perp D|B$ (i.e., $p(D|A, B) = p(D|B)$)?
3. Is $A \perp E|B$ (i.e., $p(A, E|B) = p(A|B)p(E|B)$)?
4. Is $A \perp E|B, C$ (i.e., $p(A, E|B, C) = p(A|B, C)p(E|B, C)$)?
5. Is $A \perp C|D$ (i.e., $p(A, C|D) = p(A|D)p(C|D)$)?

Please do write down your reasoning.

4 K-means [6 points]

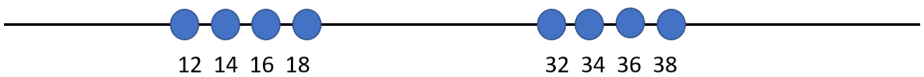


Figure 5: A one-dimensional dataset of 8 data instances

Figure 5 shows a dataset of 8 data instances, each of them is one-dimensional.

1. Please perform K-means for 10 iterations, given $K = 2$ and $c_1^{(t)} = 25$ and $c_2^{(t)} = 50$ at $t = 0$ (i.e., initialization). $c_1^{(t)}$ and $c_2^{(t)}$ represent the $K = 2$ centers at the end of the t-th iteration. Please use *squared* Euclidean distance (the same as in the slides). What will $c_1^{(10)}$ and $c_2^{(10)}$ be? [3 points]

2. Please perform K-means for 10 iterations, given $K = 2$ and $c_1^{(t)} = 20$ and $c_2^{(t)} = 60$ at $t = 0$ (i.e., initialization). $c_1^{(t)}$ and $c_2^{(t)}$ represent the $K = 2$ centers at the end of the t-th iteration. Please use *squared* Euclidean distance (the same as in the slides). What will $c_1^{(10)}$ and $c_2^{(10)}$ be? [3 points]

Note: If no points are assigned to a center, e.g., if no points are assigned to $c_1^{(t)}$, then $c_1^{(t+1)} = c_1^{(t)}$.

Please do write down your derivation/calculation.

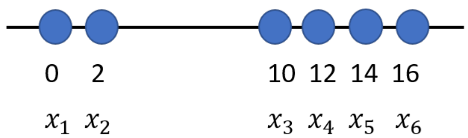


Figure 6: A one-dimensional data set of 6 data instances

5 Gaussian mixture models [14 points]

Figure 6 shows a data set of 6 data instances, each of them is one-dimensional. In the following, you will practice some steps of constructing a Gaussian mixture model with two components (i.e., 1 & 2) to model the distribution of 6 data points.

Note that to construct a Gaussian mixture model, you are to estimate

- μ_1, σ_1, such that $p(X = x|Y = 1) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp -\frac{(x-\mu_1)^2}{2{\sigma_1}^2}$,
- μ_2, σ_2, such that $p(X = x|Y = 2) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp -\frac{(x-\mu_2)^2}{2{\sigma_2}^2}$,
- λ, such that $p(Y = 1) = \lambda$,

so that together you have the GMM distribution

$$p(X = x) = p(Y = 1)p(X = x|Y = 1) + p(Y = 2)p(X = x|Y = 2).$$

In the following, let us denote by θ all the parameters $\{\mu_1, \sigma_1, \mu_2, \sigma_2, \lambda\}$.

Like K-means, GMM also needs initialization. Let

- $\mu_1^{(0)} = 4$, $\sigma_1^{(0)} = 4$;
- $\mu_2^{(0)} = 10$, $\sigma_2^{(0)} = 4$;
- $\lambda^{(0)} = 0.5$

be the initialization and let $\theta^{(0)}$ denote $\{\mu_1^{(0)}, \sigma_1^{(0)}, \mu_2^{(0)}, \sigma_2^{(0)}, \lambda^{(0)}\}$. Here, "(0)" means the parameters at the 0-th iteration.

1. **E-steps [6 points]**: Please compute $p(Y = 1|x_i; \theta^{(0)})$ for each data point in $\{x_i\}_{i=1}^6$ (a numerical value).

$$p(Y = 1|x_i; \theta^{(0)}) = \frac{p(Y = 1, x_i; \theta^{(0)})}{p(x_i; \theta^{(0)})}$$

$$= \frac{p(Y = 1; \theta^{(0)})p(x_i|Y = 1; \theta^{(0)})}{p(Y = 1; \theta^{(0)})p(x_i|Y = 1; \theta^{(0)}) + p(Y = 2; \theta^{(0)})p(x_i|Y = 2; \theta^{(0)})}$$

2. **M-steps [8 points]**: Now based on the values $p(Y = 1|x_i; \theta^{(0)})$ for $i \in \{1, \cdots, 6\}$ that you computed earlier, please compute the updated parameters $\theta^{(1)} = \{\mu_1^{(1)}, \sigma_1^{(1)}, \mu_2^{(1)}, \sigma_2^{(1)}, \lambda^{(1)}\}$.

$$\mu_k^{(1)} = \frac{\sum_i p(Y = k|x_i; \theta^{(0)}) \times x_i}{\sum_i p(Y = k|x_i; \theta^{(0)})},$$
$$\sigma_k^{(1)} = \sqrt{\frac{\sum_i p(Y = k|x_i; \theta^{(0)}) \times (x_i - \mu_k^{(1)})^2}{\sum_i p(Y = k|x_i; \theta^{(0)})}},$$
$$\lambda^{(1)} = \frac{\sum_i p(Y = 1|x_i; \theta^{(0)})}{6},$$

where $k \in \{1, 2\}$ is the index of each component. Please compute and write down the values of $\mu_1^{(1)}, \sigma_1^{(1)}, \mu_2^{(1)}, \sigma_2^{(1)}, \lambda^{(1)}$.

Please do write down your derivation/calculation.