IS5152 Data-driven Decision Making
SEMESTER 2 2023-2024
Assignment 3

---

1. (10 points) Producto Corporation has experienced a substantial increase in sales during the past few years. In order to keep up with demand it appears necessary to expand production lines. Of course, Producto could decide to do nothing to keep up with the anticipate demand, simply keeping up existing facilities with minimal maintenance. They could also modernize the existing facilities, contract some of the work to a subcontractor, or build a new additional plan.

   The four potential strategies will produce different profits in conjunction with any of three perceived states in the future: declining demand (D_1), no change in demand (D_2), and increasing demand (D_3). Profits for each strategy under each state have been estimated, as well as probabilities for occurrences of each state. These estimates are given in the following table:

	Type of future demand		
Strategy	D_1	D_2	D_3
Maintenance	300,000	25,000	-100,000
Modernization	200,000	10,000	-50,000
Subcontracting	-25,000	0	100,000
New plant	-500,000	-200,000	500,000
$P(D_i)$	0.15	0.60	0.25

   (a) What is the best strategy to maximize the potential profit?

   - Maintenance, expected return = 0.15(300,000) + 0.60(25,000) + 0.25(-100,000) = 35,000
   - Modernization, expected return = 0.15(200,000) + 0.60(10,000) + 0.25(-50,000) = 23,500
   - Subcontracting, expected return = 0.15(-25,000) + 0.60(0) + 0.25(100,000) = 21,250
   - New plant, expected return = 0.15(-500,000) + 0.60(-200,000) + 0.25(500,000) = -70,000

   Best strategy: Maintenance

   (b) What is the Expected Value With Perfect Information **EVWPI**?

   - Declining demand: maintenance, expected return: 300,000
   - No change in demand: maintenance, expected return: 25,000
   - Increasing demand: new plant, expected return: 500,000

   Expected Value With Perfect Information = 0.15(300,000) + 0.60(25,000) + 0.25(500,000) = 185,000.

   (c) What is the best strategy using the minimax-regret criterion? (note: ignore the probabilities $P(D_i)$).

   Regret matrix:

	Type of future demand			
Strategy	D_1	D_2	D_3	Max regret
Maintenance	0	0	600,000	600,000
Modernization	100,000	15,000	550,000	550,000
Subcontracting	325,000	25,000	400,000	400,000
New plant	800,000	225,000	0	800,000

Using Minimax regret criterion, the best strategy is Subcontracting.

2. (10 points) You are planning to publish your first book "My life as a student". You can either publish the book yourself or through a publisher. If you decide to publish the book yourself, the initial cost for printing and promoting the book is \$25000 and each copy of the book sold will net you a profit of \$10.

   The publisher is offering you \$8000 for signing the contract and pay you a royalty of \$1 for each book sold.

   Past records show that there is a probability of 30% that an autobiographical book by a first time author will be a success where 8000 copies were sold. When the book is not successful, it would sell only 1000 copies.

   (a) Based on the information provided, what is the best course of action?

   - Publisher: Expected gain: 8000 + 0.3(8000) + 0.7(1000) = 11100
   - Self-publish: Expected gain: 0.3(80000) + 0.7(10000) - 25000 = 6000

   Best decision: Publish through a publisher, expected value with original information = 11100.

   (b) Suppose that you may hire a literary agent to give you a review of your book and a prediction of the potential success of the book. From past experience, you know that when a book is successful, the agent will predict the correct outcome 75% of the time. When the book was not successful, the agent was correct 90% of the time. How would this information affect your decision?

   Let

   - S = the book is successful
   - F = the book is a failure
   - PS = the book is predicted as successful
   - PF = the book is predicted as a failure

   Then

   - $P(PS|S) = 0.75$, $P(PF|S) = 0.25$, $P(PS|F) = 0.1$, $P(PF|F) = 0.9$
   - $P(PS \cap S) = P(S) \times P(PS|S) = 0.3 \times 0.75 = .225$
   - $P(PS \cap F) = P(F) \times P(PS|F) = 0.7 \times 0.1 = .07$
   - $P(PF \cap S) = P(S) \times P(PF|S) = 0.3 \times 0.25 = .075$
   - $P(PF \cap F) = P(F) \times P(PF|F) = 0.7 \times 0.9 = .63$
   - $P(PS) = P(PS \cap S) + P(PS \cap F) = 0.295$
   - $P(PF) = P(PF \cap S) + P(PF \cap F) = 0.705$

   Compute revised probabilities

- P(S|PS) = P(S∩PS)/P(PS) = 0.225/0.295 = 0.7627
- P(F|PS) = P(F∩PS)/P(PS) = 0.07/0.295 = 0.2373
- P(S|PF) = P(S∩PF)/P(PF) = 0.075/0.705 = 0.1064
- P(F|PF) = P(F∩PF)/P(PF) = 0.63/0.705 = 0.8936

If the literary agent is hired:

- She predicts S:
  - Let publisher publish the book: Expected return = 0.7627×8000 + 0.2373×1000 + 8000 = 14338.90
  - Self publish the book: Expected return = 0.7627×80000 + 0.2373×10000 - 25000 = 38389.00
  - Decision: Self publish.
- She predicts F:
  - Let publisher publish the book: Expected return = 0.1064×8000 + 0.8936×1000 + 8000 = 9744.80
  - Self publish the book: Expected return = 0.1064×80000 + 0.8936×10000 - 25000 = -7552.00
  - Decision: Let publisher publsih the book.

If the agent is hired, if she predicts S, self-publish the book, otherwise, publish through the publisher. Expected return when the agent is hired:

$$0.295 \times max\{14338.90, 38389.00\} + 0.705 \times max\{9744.80, -7552.00\} = 18194.89$$

Whether or not the agent should be hired, it depends on the cost.

(c) Compute the Expected Value of Perfect Information and the Expected Value of Sample Information.

- If the book is going to be a success:
  - Self publish, expeceted return = $8000 \times 10 - 25000 = 55000$
  - Let publisher publish, expected return = $8000 + 8000(1) = 16000$
  - Decision: self publish.
- If the book is going to be a failure:
  - Self publish, expeceted return = $1000 \times 10 - 25000 = -15000$
  - Let publisher publish, expected return = $8000 + 1000(1) = 9000$
  - Decision: let publisher publish.
- Expected value with perfect information: $0.3 \times 55000 + 0.7 \times 9000 = 22800$
- Expected value with original information from part (a): 11100.
- Expected value of perfect information: $22800 - 11100 = 11700$
- Expected value of sample information: $18194.89 - 11100 = 7094.89$.

3. (10 points) The following table consists of training data from an employee dataset. The attribute *department* has 3 possible values: sales, systems, marketing. The attribute *age* has 4 possible values: [21 to 25], [26 to 30], [31 to 35], [36 to 40]. The attribute *salary* has 3 possible values: $< 30K$, $[30K$ to $40K]$, $> 40K$. The target attribute *status* is binary-valued: junior or senior.

department	age	salary	status
sales	[31 to 35]	$> 40K$	senior
sales	[26 to 30]	$[30K - 40K]$	junior
systems	[21 to 25]	$< 30K$	junior
systems	[36 to 40]	$> 40K$	senior
systems	[26 to 30]	$> 40K$	junior
systems	[31 to 35]	$[30K - 40K]$	junior
marketing	[31 to 35]	$[30K - 40K]$	senior
marketing	[36 to 40]	$[30K - 40K]$	senior
marketing	[26 to 30]	$> 40K$	senior
marketing	[21 to 25]	$[30K - 40K]$	junior

(a) What would be the naive Bayes prediction for the status of an employee with the attribute values: (marketing,[31 to 35], $> 40K$)?

- P(senior) = 5/10, P(junior) = 5/10
- P(marketing|senior) = 3/5
- P([31,35]|senior) = 2/5
- P($> 40K$|senior) = 3/5
- P(senior)*P(marketing|senior)*P([31,35]|senior)*P($> 40K$|senior) = 9/125 = 0.072
- P(marketing|junior) = 1/5
- P([31,35]|junior) = 1/5
- P($> 40K$|junior) = 1/5
- P(junior)*P(marketing|junior)*P([31,35]|junior)*P($> 40K$|junior) = 1/250 = 0.004
- Predict: senior

(b) Laplace smoothing is applied to the data with parameter k set to 3. What is the predicted status for an employee with attribute values: (sales,[36 to 40],$< 30K$)?

- original: P(sales|senior) = 1/5
- original: P([36-40]|senior) = 2/5
- original: P($< 30K$|senior) = 0
- original: P(sales|junior) = 1/5
- original: P([36-40]|junior) = 0
- original: P($< 30K$|junior) = 1/5
- department: 3 levels, P(sales|senior) = 4/(5 + 9) = 4/14
- department: 3 levels, P(sales|junior) = 4/(5 + 9) = 4/14
- age: 4 levels, P([36-40]|senior) = 5/(5 + 12) = 5/17
- age: 4 levels, P([36-40]|junior) = 3/(5 + 12) = 3/17
- salary: 3 levels, P($< 30K$|senior) = 3/(5 + 9) = 3/14
- salary: 3 levels, P($< 30K$|junior) = 4/(5 + 9) = 4/14
- For senior: P(senior)*P(sales|senior)*P([36-40]|senior)*P($< 30K$|senior) $= (5/10) \times (4/14) \times (5/17) \times (3/14) = 0.00900$
- For junior: P(junior)*P(sales|junior)*P([36-40]|junior)*P($< 30K$|junior) $= (5/10) \times (4/14) \times (3/17) \times (4/14) = 0.00720$
- Predict senior