

Example 1

1. A piece of _____ is on the plate.
2. Everyone enjoys eating _____.
3. You can cut _____ with a knife.
4. We make _____ from milk.

Potential Answers: "cheese," "cake," or "butter"

Steps for Example 1

- **Sentence 1:** A piece of _____ is on the plate. All three options (cheese, cake, butter) can be referred to as a piece.
- **Sentence 2:** Everyone enjoys eating _____. All are edible and generally enjoyed, though cake might be more universally associated with enjoyment.
- **Sentence 3:** You can cut _____ with a knife. All can be cut with a knife
- **Sentence 4:** We make _____ from milk. Cheese and butter are dairy products made from milk, whereas cake is not (it may contain milk but is not made from milk in the same direct way).

Conclusion: Both “cheese” and “butter” fit Sentence 4, but “cake” does not. Since “cheese” is more commonly referred to as a “piece” on a plate and cut with a knife in everyday contexts, **cheese** is the most likely answer.

For Example, 1	Cheese	Cake	Butter
Sentence 1	1	1	1
Sentence 2	1	1	0
Sentence 3	1	1	1
Sentence 4	1	0	1

From the distributional Hypothesis, we can conclude that **Cheese** would be the potential answer for Example 1

ANS: Cheese

Example 2

1. The _____ is parked in the driveway.
2. He bought a new _____ for his birthday.
3. _____ can drive really fast.
4. People often wash their _____ on the weekends.

Potential Answers: "car," "truck," or "motorcycle"

Steps for Example 2

- **Sentence 1:** The _____ is parked in the driveway. All three (car, truck, motorcycle) can be parked in a driveway.
- **Sentence 2:** He bought a new _____ for his birthday. All are plausible purchases for a birthday, so cars and trucks might be more common due to cost and practicality.
- **Sentence 3:** _____ can drive really fast. All can be driven fast, but motorcycles are particularly associated with speed.
- **Sentence 4:** People often wash their _____ on the weekends. Cars and trucks are frequently washed in driveways, and motorcycles can be too, so car washing can be a more common weekend activity.

Conclusion: All options fit, but “car” is the most general and commonly associated with driveway

Sentence	Car	Truck	Motorcycle
Sentence 1	1	1	1
Sentence 2	1	1	1
Sentence 3	1	1	1
Sentence 4	1	1	1

From the distributional Hypothesis, we can clearly conclude that **Car** would be the potential answer for Example 2.

ANS: Car

Example 3

1. I read an interesting _____ last night.
2. Many people enjoy a good _____ before bed.
3. _____ often has chapters and a cover.
4. You can borrow a _____ from the library.

Potential Answers: "book," "novel," or "story" – these terms appear in contexts related to reading and library usage.

Steps for Example 2

- Sentence 1: I read an interesting _____ last night. All fit: you can read a book, novel, or story.
- Sentence 2: Many people enjoy a good _____ before bed. Books and novels are commonly read before bed stories are less typical unless referring to a short story within a book.
- Sentence 3: _____ often has chapters and a cover. Books and novels typically have chapters and covers a story (as a standalone concept) do not.
- Sentence 4: You can borrow a _____ from the library. Libraries lend books and novels story is not a typical library item.

Conclusion: "Book" and "novel" fit best, but "book" is more general and encompasses novels. Book is the most consistent across all sentences.

Sentence	Book	Novel	Story
Sentence 1	1	1	1
Sentence 2	1	1	0
Sentence 3	1	1	0
Sentence 4	1	1	0

From the distributional Hypothesis, we can conclude that Book would be the potential answer for Example 3.

Part II

Python Notebook is submitted in the form of an ipynb notebook separately

Summary: I used TFIDF vectorization on the data tales dataset after performing some simple preprocessing (lowercasing, punctuation removal) for this work. Using five-fold cross-validation, I trained and assessed four classifiers: Random Forest, SVM, Naive Bayes, and Logistic Regression. With an accuracy of about 85.3%, the Support Vector Machine outperformed the others. Furthermore, I used Logistic Regression to perform a Leave-One-Plot-Out cross-validation, which produced an accuracy of about 70.0%. This suggests that while logistic regression generalizes quite well across unseen plots, support vector machines (SVM) perform best under typical CV circumstances.