Machine Learning Self-Study Worksheet

Worksheet: 2025002

Level: Easy
Topics Covered: Supervised Learning

Part 1: Conceptual Questions

<table>
<tr><th>Q. No.</th><th>Question</th><th>Points</th></tr>
<tr><td>1.</td><td>Classification vs. Regression
• You are building a machine learning model that estimates the selling price of a house using features like square footage, location, number of bedrooms, and age of the property.
→ Is this a classification or regression problem? Explain why.
• You are designing an email filter that identifies whether a message is "spam" or "not spam" based on the words used, sender address, and frequency of certain phrases.
→ Is this a classification or regression problem? Justify your answer.</td><td>5</td></tr>
<tr><td>2.</td><td>Define supervised classification and complete the following table about the phases of a classification algorithm:
<table><tr><th>Phase</th><th>Input</th><th>Process</th><th>Output</th></tr><tr><td>Training Phase</td><td></td><td></td><td></td></tr><tr><td>Testing Phase</td><td></td><td></td><td></td></tr><tr><td>Prediction Phase</td><td></td><td></td><td></td></tr></table></td><td>5</td></tr>
</table>

Part 2: Numerical Questions

Q. No.	Question	Points
3.	You are given the multiple linear regression equation: $Y = 2X_1 + 4X_2 + 3$ Find the value of Y for: • $X_1 = 1,\ X_2 = 2$ • $X_1 = 3,\ X_2 = 1$	5
4.	Given data points: • Class A: (1, 1), (2, 1), (2, 2) • Class B: (6, 6), (7, 7), (8, 6) Use **Euclidean distance** and determine the class of (4, 3) using the **3-nearest neighbors**.	5

Part 3: Coding Exercise

Q. No.	**Complete & Run the Code** (Use your own Python IDE or Google Colab.) Complete the KNN classification code for k = 3 and predict the class of a new point:	Points

5.	from sklearn.linear_model import _____
	from sklearn.model_selection import train_test_split
	from sklearn.metrics import accuracy_score
	
	# Dataset
	X = [[2], [4], [6], [8], [10], [12]]
	y = [0, 0, 0, 1, 1, 1]
	
	# Train-test split
	X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33)
	
	# Model
	model = LogisticRegression()
	model._____(X_train, y_train)
	
	# Predict
	y_pred = model._____(X_test)
	print("Accuracy:", accuracy_score(y_test, y_pred))
		5

Notes & Disclaimer

- For quick notes or queries, visit: https://www.instagram.com/kparse_code/ or https://www.youtube.com/channel/UCGuhk1P1OA0X_7Ar2-xadEw .

- **Related Resources**:
  - Google Colab – to run Python code online without installing anything
  - StatQuest: https://www.youtube.com/@statquest/playlists
  - 3Blue1Brown: https://www.youtube.com/@3blue1brown/playlists
  - Numerical: https://www.youtube.com/@MaheshHuddar/playlists

- **Want more worksheets like this?**
  Try using **AI prompt** like:
  "Create a beginner-level machine learning worksheet with 2 conceptual questions(with sub-questions), 2 numerical, and 1 code-completion exercise on [topic(s)]"

- **How to Use This Worksheet**:
  - These worksheets are designed for **self-study and concept reinforcement**.
  - Try to solve without looking up the answers first.
  - Use a notebook or Python IDE to experiment — code is the best teacher!
  - Feel free to modify or extend questions as your understanding grows.

- **Disclaimer**:
  This worksheet is for self study & **educational purposes only**. Accuracy of content may vary depending on updates to libraries or definitions. Always refer to official documentation or textbooks for exam-level prep.