

✓ Task 11: SVM – Breast Cancer Classification

Dataset:

- Primary: Sklearn Breast Cancer – `load_breast_cancer()`
- Alternative: Kaggle Breast Cancer dataset

Tools:

- Python
- Scikit-learn
- Matplotlib
- Alternatives: Weka, R caret

Hints / Mini Guide:

1. Load dataset and inspect features and labels distribution.
2. Apply StandardScaler to normalize feature values.
3. Split data into train-test sets.
4. Train baseline SVM with linear kernel and check performance.
5. Train SVM with RBF kernel and compare accuracy.
6. Use GridSearchCV to tune C and gamma values.
7. Evaluate best model using confusion matrix and classification report.
8. Plot ROC curve and calculate AUC score.
9. Save tuned model pipeline (scaler + svm) for reuse.

Deliverables:

- Notebook
- ROC curve + AUC report
- Saved model file

Final Outcome:

- Intern learns kernel-based classification and tuning.

Interview Questions Related To Above Task:

- What is margin in SVM?
- Difference between linear and RBF kernel?
- What is C parameter?
- What is gamma?
- Why scaling required for SVM?

Task Submission Guidelines

-  **Time Window:**

You can complete the task anytime between 10:00 AM to 10:00 PM on the given day. Submission link closes at 10:00 PM

-  **Self-Research Allowed:**

You are free to explore, Google, or refer to tutorials to understand concepts and complete the task effectively.

-  **Debug Yourself:**

Try to resolve all errors by yourself. This helps you learn problem-solving and ensures you don't face the same issues in future tasks.

-  **No Paid Tools:**

If the task involves any paid software/tools, do not purchase anything. Just learn the process or find free alternatives.

-  **GitHub Submission:**

Create a new GitHub repository for each task.

Add everything you used for the task — code, datasets, screenshots (if any), and a short README.md explaining what you did.

Submit Here:

After completing the task, paste your GitHub repo link and submit it using the link below:

-  [\[Submission Link\]](#)

