Here’s the full set of answers step by step based on your requirements:

**1. Choose a Binary Classification Dataset**

We used the **Breast Cancer dataset** from sklearn.datasets.

* It has **569 samples**, **30 features**, and a **binary target**:
  + 0 = malignant
  + 1 = benign

**2. Train/Test Split and Standardize Features**

* We split the dataset into **80% training** and **20% testing** using train\_test\_split.
* Applied **StandardScaler** to normalize features so they have zero mean and unit variance.

**3. Fit a Logistic Regression Model**

* Used **LogisticRegression(max\_iter=1000)** from sklearn.
* Trained on scaled training data.
* Predictions:
  + y\_pred for class labels
  + y\_prob for probabilities

**4. Evaluate with Confusion Matrix, Precision, Recall, ROC-AUC**

**Confusion Matrix (Threshold = 0.5)**

|  | **Pred 0** | **Pred 1** |
| --- | --- | --- |
| **Actual 0** | 39 | 4 |
| **Actual 1** | 2 | 69 |

* **Precision:** 0.9452
* **Recall:** 0.9718
* **ROC-AUC:** 0.9986

**5. Tune Threshold (Example: 0.4)**

Changing threshold from **0.5 → 0.4**:

* New confusion matrix:

|  | **Pred 0** | **Pred 1** |
| --- | --- | --- |
| **Actual 0** | 38 | 5 |
| **Actual 1** | 1 | 70 |

* **Precision (0.4):** 0.9333
* **Recall (0.4):** 0.9859

👉 **Observation:** Lowering threshold improves recall but slightly decreases precision.

**Sigmoid Function Explanation**

The **sigmoid function** converts a linear combination of inputs into a probability between 0 and 1:

σ(z)=11+e−z\sigma(z) = \frac{1}{1 + e^{-z}}

Where:

* z=w⋅x+bz = w \cdot x + b
* Output ∈(0,1)\in (0,1) is interpreted as **probability of class 1**.

In logistic regression:

* If σ(z)≥threshold\sigma(z) \ge \text{threshold}, predict class 1; else class 0.
* Default threshold = 0.5, but you can adjust it to control **precision vs recall**.

✅ I’ve included all these details in the PDF you downloaded earlier.

Do you want me to **give you the Python code in text form too**, so you can run it yourself? Or should I **add a ROC curve image + threshold vs metrics graph and regenerate the PDF**?