



CSST102 - Basic Machine Learning	
Macatangay, Shaine Carla P.	BSCS 3A

Hour 1 – Setup & Dataset Exploration (Mini-task)

What is the input (features)?

- For the Iris dataset, the input is the data we use to make predictions so the features are **sepal length, sepal width, petal length, and petal width**. These are the measurements of the flowers.

What is the output (label)?

- The output is what we want to predict. For the Iris dataset, the label is the flower species Iris Setosa, Iris Versicolor, or Iris Virginica.

Is this supervised or unsupervised learning?

- This is supervised learning because we already know the correct answers (the flower species) and we use them to train the model.

Hour 2 – Train-Test Split & Baseline Model (Mini-task)

Compute model accuracy:

- After training the Logistic Regression model on the Iris dataset, the accuracy is usually around 95% or higher. The exact value depends on the random split, but normally it's between 94% and 98%.

Hour 3 – Evaluation & Reflection

What would happen if the dataset had missing or wrong values?

- If there are missing or wrong values, the model will not learn correctly and can make mistakes in its predictions. This means the accuracy will go down, and sometimes the model might not even train properly.

How does this relate to real-world ML applications?

- In the real world, data is not always perfect because it can have errors, missing parts, or even fake information. That is why data cleaning and preprocessing are very important steps before training a model.

Conclusion:

We used supervised learning because the dataset has inputs and outputs with correct answers. The model we used is Logistic Regression for classification, which gave high accuracy on the Iris dataset. One challenge that might affect the model is overfitting, where the model learns too much from the training data and does not work well on new data. Another challenge is bad data, like missing values or incorrect information, which can confuse the model and reduce accuracy. In real-world situations, preparing the dataset carefully and avoiding overfitting are important steps to make the model reliable and useful.