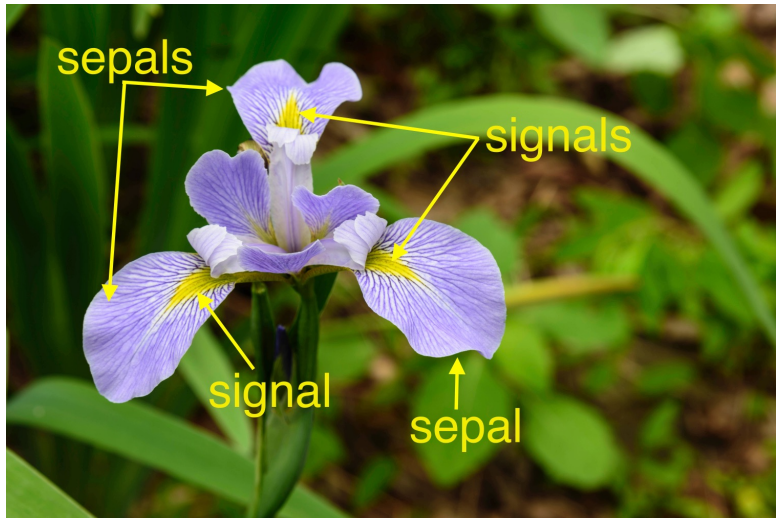
Linear Classifier Acitivity

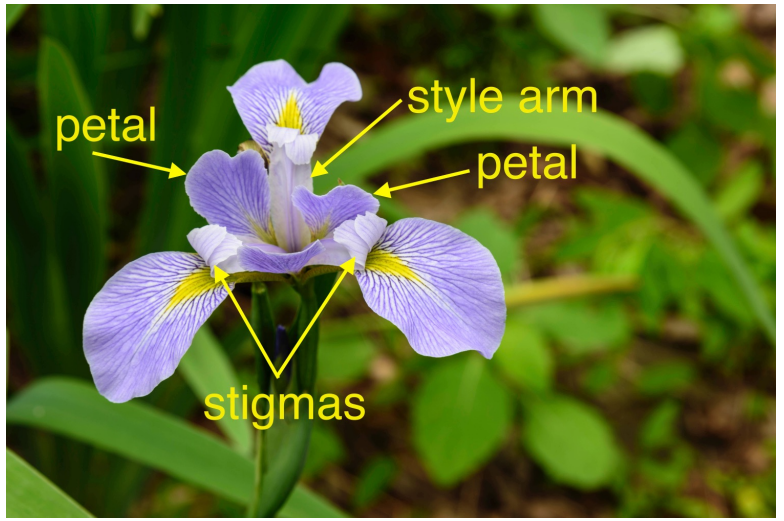
SYDE 599 Deep Learning F23

September 19, 2023

Logistic regression activity

- Fisher's Iris Dataset:
  - 3 species of iris (setosa, versicolor, vinginica)
  - 4 measurements for each flower (petal and sepal length and width)
  - 50 samples per species

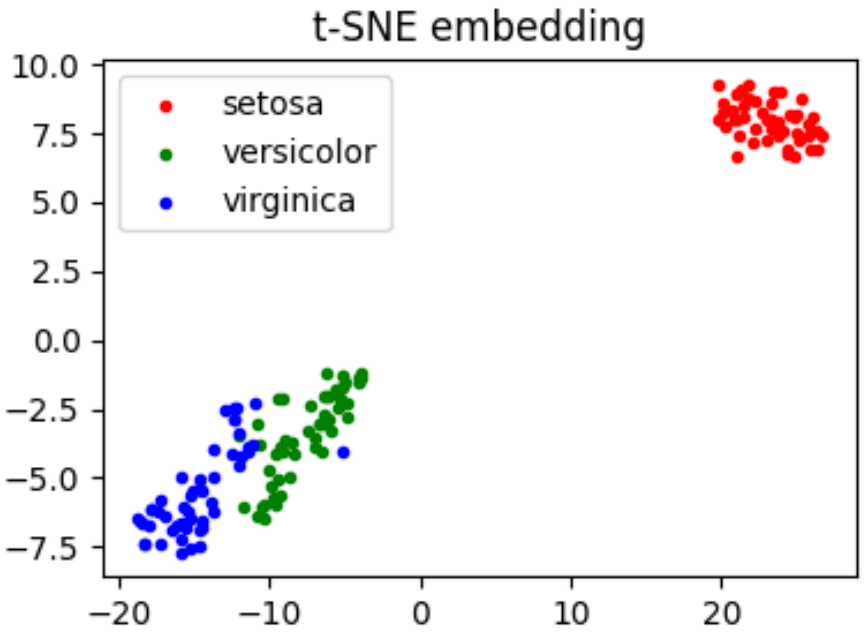




https://elizabethswildflowerblog.com/2017/05/26/iris-flowers-terminology-and-structure/

Logistic regression activity

- Here is a dimension reduction of the samples
- The setosas are distinct but the versicolor and virginica are closer to each other



Logistic regression activity

- Use logistic regression to estimate the probability of a set of measurements being versicolor and virginica
- Use the dataset on Learn and sklearn.linear_model. LogisticRegression
- Plot histograms of the predicted probabilities for samples in each category for both the training and validation datasets
- Code to start with:

```
import pickle
with open('logistic-regression-data.pkl', 'rb') as f:
    data = pickle.load(f)
```