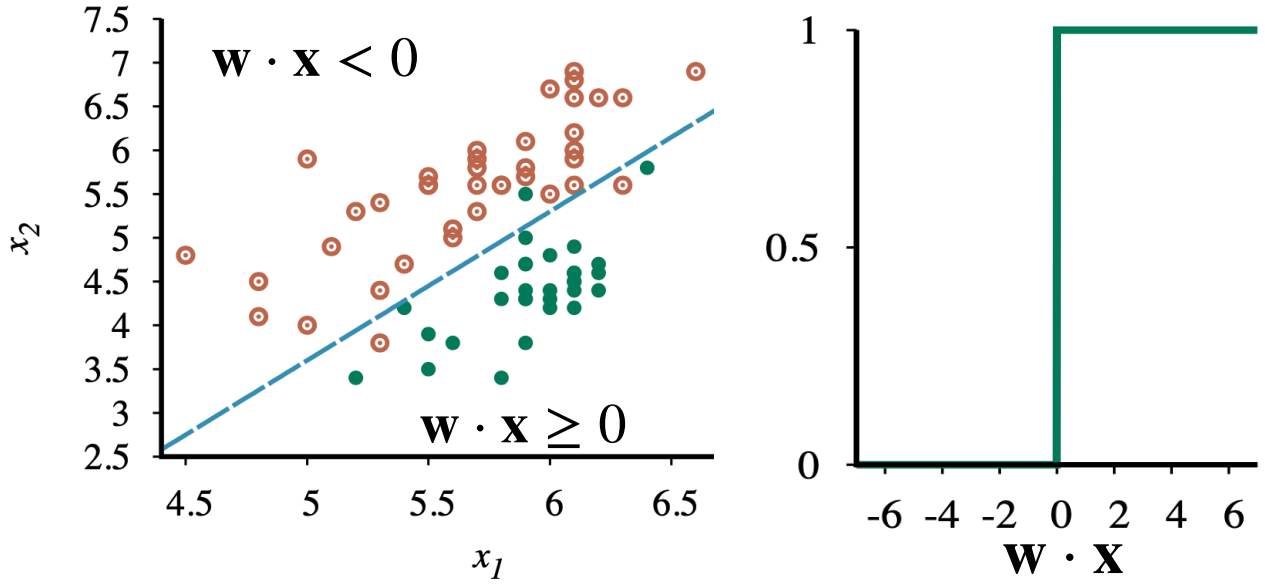
Python: Classification

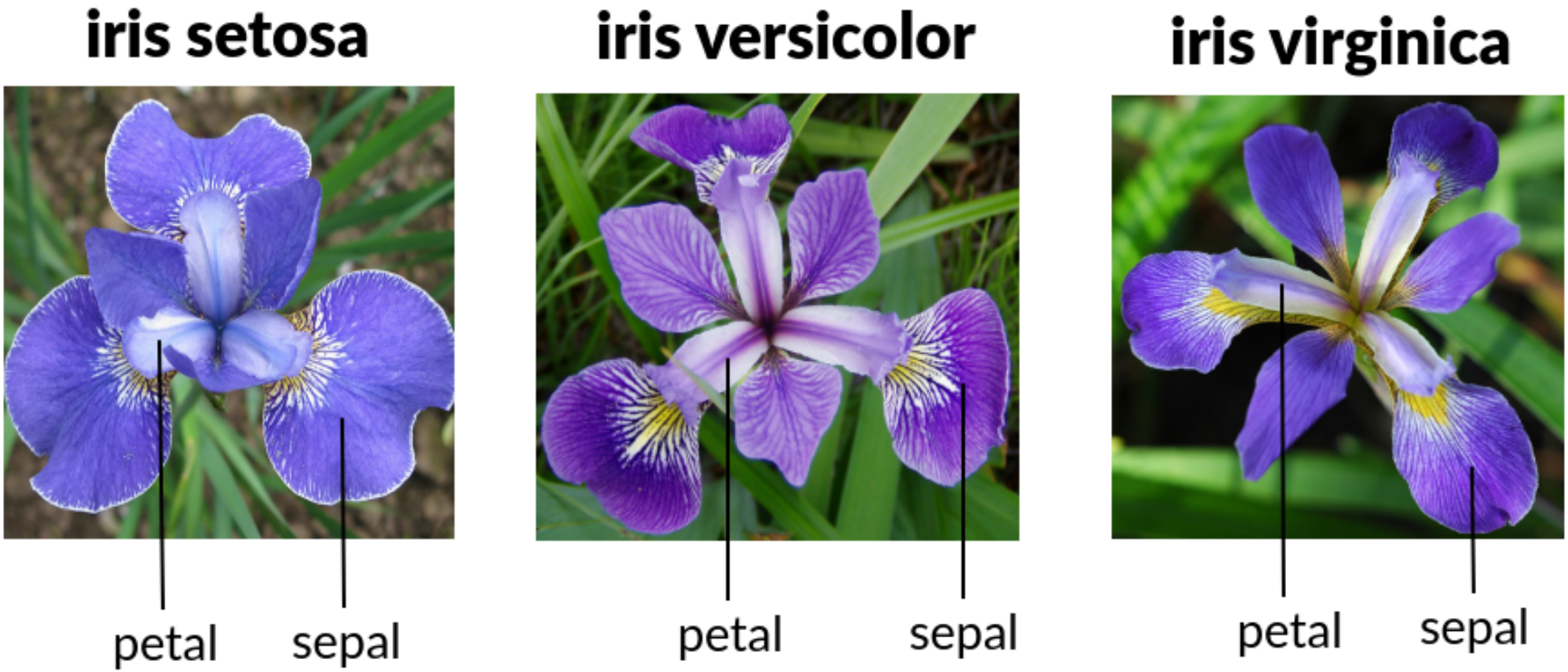
Python: Perceptron Learning Rule

- Given data points of two classes: 0/1, learn a classification hypothesis h

$$h_{\mathbf{w}}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



Iris Flowers Classification



Source: https://editor.analyticsvidhya.com/uploads/51518iris img1.png

Normalization

- **Min-max normalization**

$$x'_{j,i} = \frac{x_{j,i} - min_i}{max_i - min_i}$$

  - max_i is the maximum of the values in each dimension
  - min_i is the minimum of the values in each dimension
  - The new value $x'_{j,i}$ is in [0,1]

Normalization

- **Z-score normalization (Standardization)**

$$x'_{j,i} = \frac{x_{j,i} - \mu_i}{\sigma_i}$$

  - μ_i is the mean of the values in each dimension
  - σ_i is the standard deviation of the values in each dimension

P2: Programming Assignment (Due: 5/27)

- Task: Binary classification for Disease Prediction
- Dataset
  - Target (two labels): 0/1
  - Input: 17 attributes/features
  - Total: 550 records

- Age
- RestingBP: 靜息血壓
- Cholesterol: 毫克/分升血清膽固醇
- FastingBS: 0/1, 空腹血糖 > 120毫克/分升
- MaxHR: 達到最大心率
- Oldpeak: 運動相對於休息引起的ST段下降
- Male: True/False
- ChestPainType_ASY: True/False
- ChestPainType_ATA: True/False
- ChestPainType_NAP: True/False
- ChestPainType_TA: True/False
- RestingECG_LVH: True/False
- RestingECG_ST: True/False
- ExercisesAngina: True/False, 運動誘導的心絞痛
- ST_Slope_Down: True/False, 運動高峰ST段的斜率向下
- ST_Slope_Flat: True/False, 運動高峰ST段的斜率平坦
- ST_Slope_Up: True/False, 運動高峰ST段的斜率向上

P2: Programming Assignment (Due: 5/27)

- JUST UPLOAD the file: **submission.py**
- Task:
  - Command: 打在terminal
    - python3 submission.py disease_60.csv test01.csv answer01.csv
  - **Training data**: ./task/disease_60.csv
  - **Testing data 1(80%)**: ./task/test01.csv
    - Label: ./task/answer01.csv
  - **(On-line) Testing data 2 (20%):** ./task/test02.csv
    - Label: ./task/answer02.csv