Supervised Learning Dataset Overview
Meeting 8/8

Bo Han, Chen

National Yang Ming Chiao Tung University, Taiwan
bhchen312551074.cs12@nycu.edu.tw

August 8, 2023

Public Dataset

- public image dataset
  - Oxford 102 Flower
  - source: Visual Gepmetry Group website
  - 8189 images consist of 12 categories
  - classifier: CNN based network (ResNet), XGBoost
- public non-image dataset
  - NSL-KDD
  - source: UNB website
  - 41 features for each instance, 4 classes of attack
  - classifier: KNN, SVM

Self-Made Dataset

- MLB Team Performance Prediction
    - can statistics help teams win the game?
    - source: Baseball Reference website
    - features: the statistics of each team from 2013-2022
    - three main field: batting, pitching, fielding
    - classification: 50%+ Win-Loss percentage in the regular season or not
    - classifier: random forest, XGBoost

Thanks for Listening

Q & A