



CIS 635 Knowledge Discovery & Data Mining

Predictive modeling: Classification



Regression vs Classification

- Main difference is the target or response variable (y)



Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)



Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- **Regression examples:**
 - Diabetes scores
 - Healthcare cost



Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
 - Diabetes scores
 - Healthcare cost

- **Classification examples:**
 - **Character recognition (10 classes)**
 - Yes/No (or Binary) questions:
 - Accept/reject loan application
 - Positive vs negative sentiment
 - Dog vs cat (2 classes), still binary class



Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
 - Diabetes scores
 - Healthcare cost

- **Classification examples:**
 - Character recognition (10 classes)
 - **Yes/No (or Binary) questions:**
 - Accept/reject loan application
 - Positive vs negative sentiment
 - Dog vs cat (2 classes), still binary class



Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
 - Diabetes scores
 - Healthcare cost

- **Classification examples:**
 - Character recognition (10 classes)
 - Yes/No (or Binary) questions:
 - Accept/reject loan application
 - Positive vs negative sentiment
 - **Dog vs cat (2 classes), still binary class**



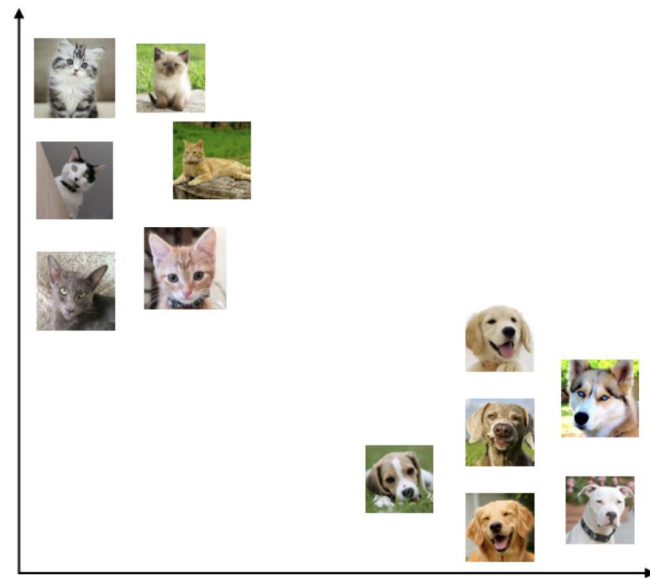
Regression vs Classification

- Main difference is the target or response variable (y)
- One predicts real/floating-point values whereas the other predicts categorical values (predefined set)
- Regression examples:
 - Diabetes scores
 - Healthcare cost
- We have learned about regression (not complete yet; will continue ..)

- Classification examples:
 - Character recognition (10 classes)
 - Yes/No (or Binary) questions:
 - Accept/reject loan application
 - Positive vs negative sentiment
 - Dog vs cat (2 classes), still binary class
- We will start our classification predictive modeling journey today

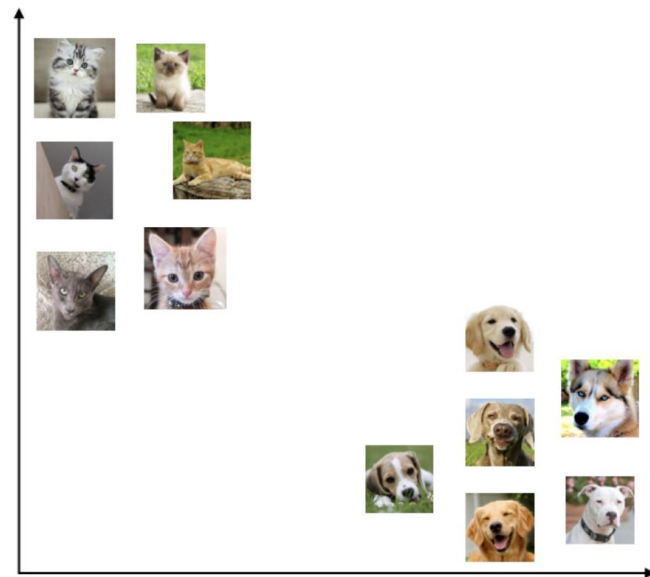
Classification

- Here we are seeing some examples of **Dog** and **Cat** images



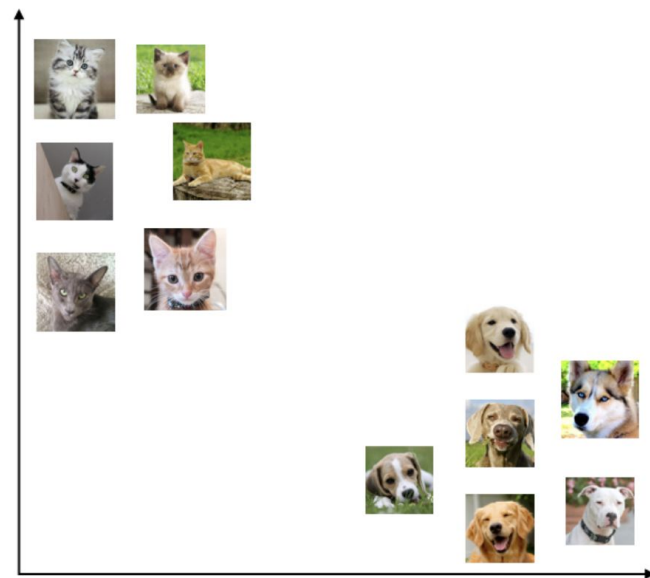
Classification

- Here we are seeing some examples of **Dog** and **Cat** images
- Both animals have features such as size, color, weight, etc.



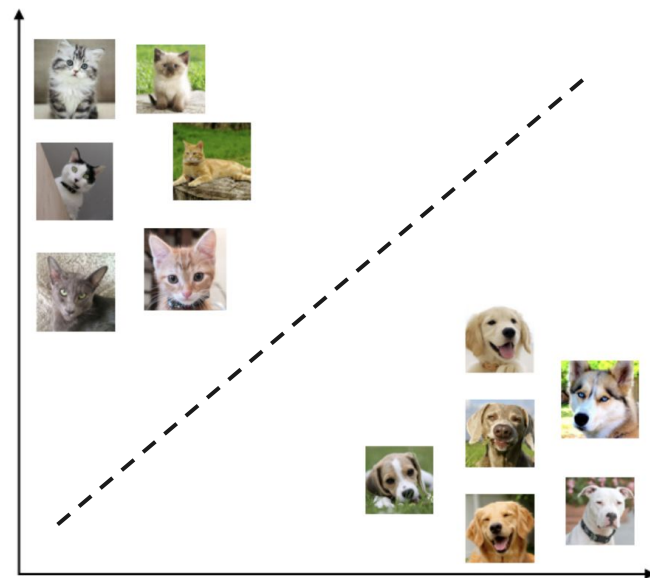
Classification

- Here we are seeing some examples of **Dog** and **Cat** images
- Both animals have features such as size, color, weight, etc.
- **We are just plotting their images on 2D plane for easy understanding.**

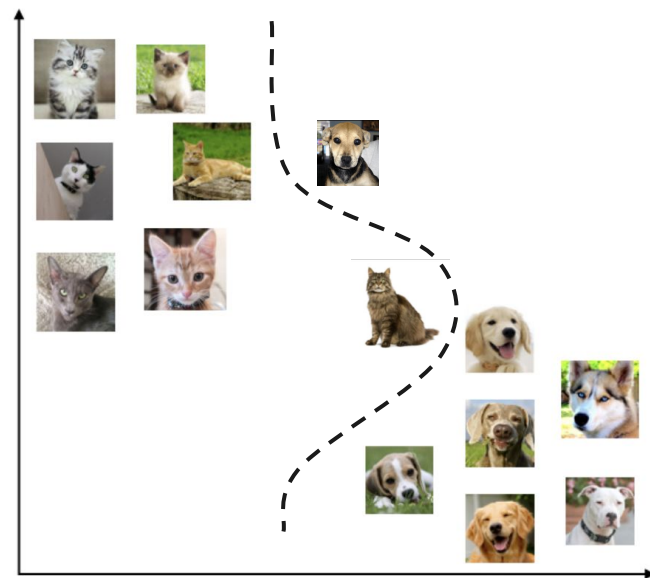


Classification

- Here we are seeing some examples of **Dog** and **Cat** images
- Both animals have features such as size, color, weight, etc.
- We are just plotting their images on 2D plane for easy understanding.
- We can separate the instances by simply using a liner straight line (a **linear classifier**):
 - On left top we have **Cats**
 - And right bottom we have **Dogs**

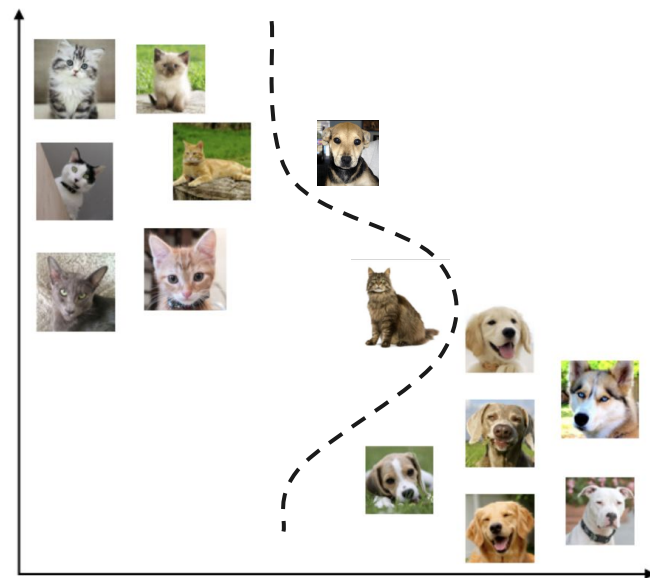


- A simple Linear (2D) classifier doesn't work for this setup



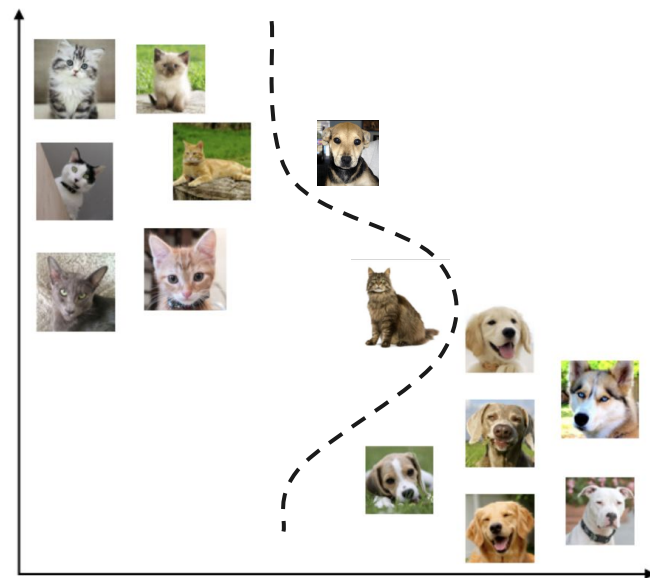
Classification

- A simple Linear (2D) classifier doesn't work for this setup
- **We require**
 - Either go to higher dimensions, or
 - Chose a non-linear classifier



Classification

- A simple Linear (2D) classifier doesn't work for this setup
- **We require**
 - Either go to higher dimensions, or
 - **Chose a non-linear classifier**





Classification Models

- Logistic Regression
- Random Forest Classifier
- Support Vector Machines (SVMs)
- Boosting Classifiers
- Naive Bayes

Logistic Regression

- Probabilistic classifier

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

- Sigmoid function

