

Basic Machine Learning

Rong Ge

Example: Dogs vs Cats

<https://www.kaggle.com/c/dogs-vs-cats>



Input: Many images with labels (cat or dog)

Goal: Given new image, decide cat or dog.

Example: Netflix



Input: Movie ratings from users
Goal: Recommend new movies for users

Example: Community Finding



Input: List of friends
Goal: Find “communities”

Supervised vs Unsupervised

Input: (data, label)
Output: Function f
Hope: $f(\text{data}) = \text{label}$

Supervised



Input: data
Output: “structure”
Hope: explain and predict

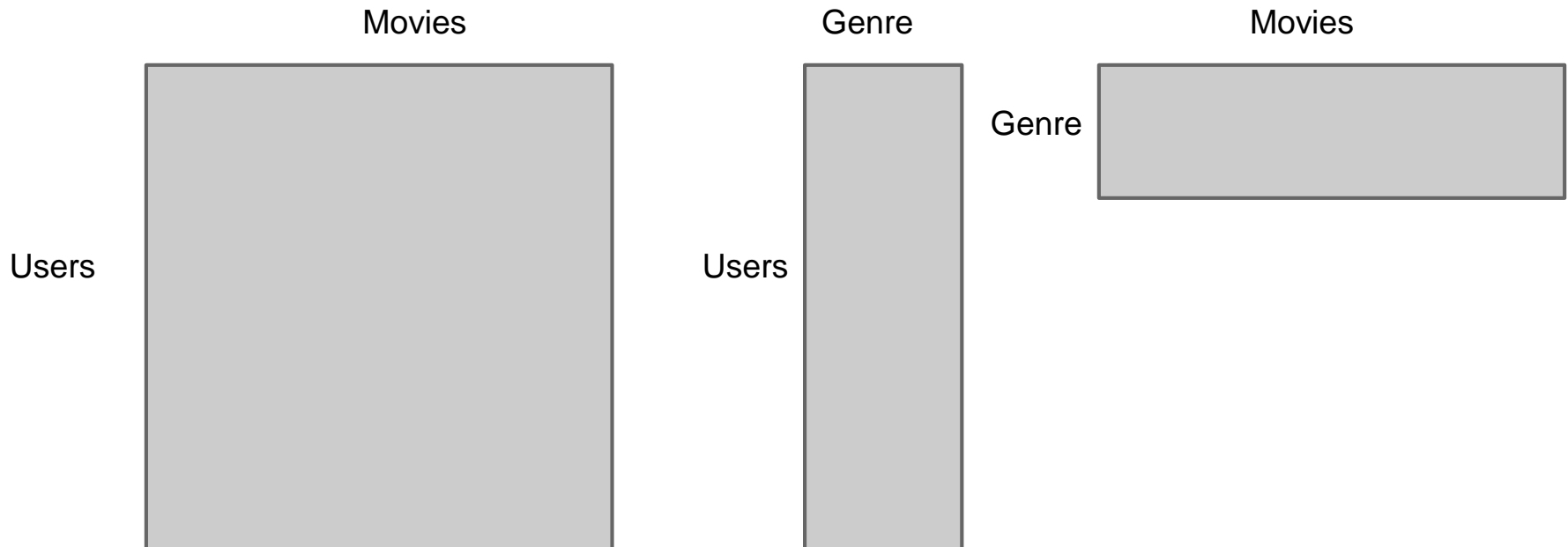
Unsupervised



Example of Structure



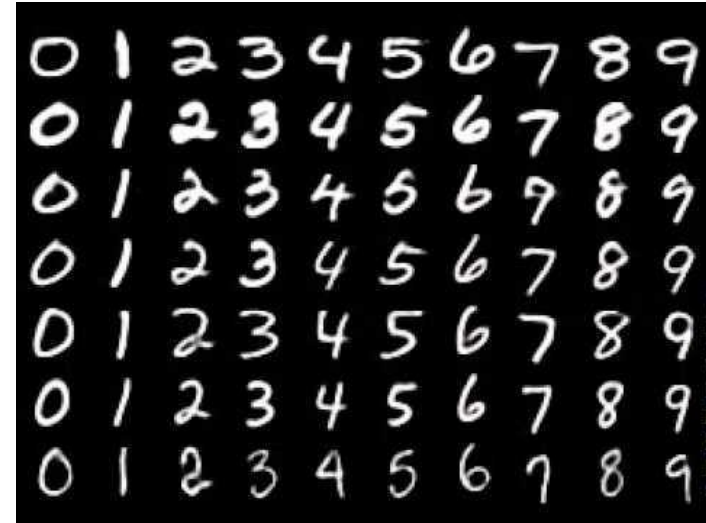
- Movies have different genres.
- Users like different genres.
- Learning ~ Find the genres of movies and users
- To recommend
 - determine what genres the user likes
 - recommend a good movie in that genre.



Connection



MNIST: Given images recognize digits.



Unsupervised: Given images, cluster similar images

Unsupervised Learning \Rightarrow features for Supervised Learning

How do we know machines have learned?

(General approach: Give an exam)

Human Learning, Machine Learning

Take a course

Understand the course
material

Take an exam

Goal of Exam: If the students
understood the material, do
well in exam.

Get training examples

Learn a *hypothesis*

- maps input to labels

Test on new examples

Goal of Test: If the
hypothesis is **good**, do well
in test samples.

What is a good hypothesis?

Ways to fail an exam

Do not understand examples given in class
(Make many mistakes on training samples)

Only memorize the basic examples

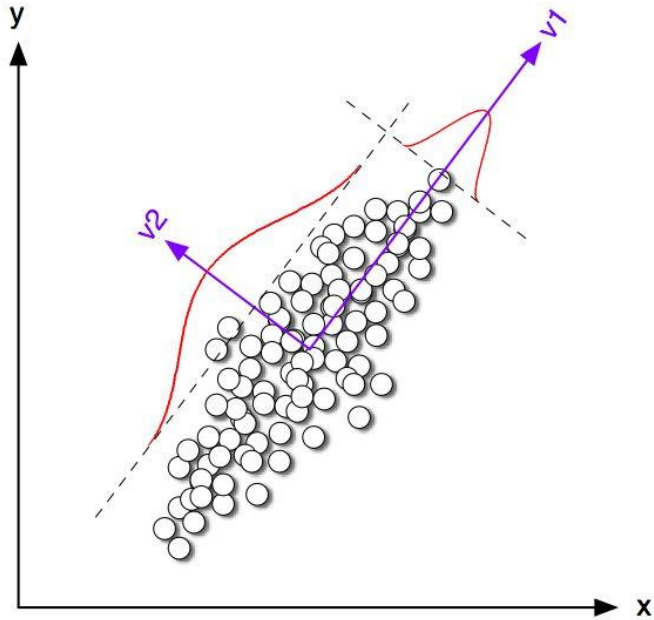
(Come up with a **complicated** hypothesis)

Good hypothesis = **Simple** + Do well on training

What we will see

PCA

Find basic directions



Least Squares/
Gradient Descent

