Lecture 1: Course logistics, homework 0

STATS 202: Data mining and analysis

Sergio Bacallado

September 23, 2013

Syllabus

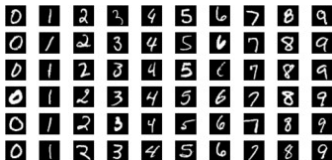- **Videos:** Every lecture will be recorded by SCPD.

Syllabus

- **Videos:** Every lecture will be recorded by SCPD.

- **Email policy:** Please email the course staff mailing list: stats202-aut1314-staff@lists.stanford.edu.

Syllabus

- **Videos:** Every lecture will be recorded by SCPD.

- **Email policy:** Please email the course staff mailing list: stats202-aut1314-staff@lists.stanford.edu.

- **Class website:** stats202.stanford.edu. If you are not registered on Axess, email us your SUNet ID in order to gain access to the lectures and homework.

Prediction challenges

The MNIST dataset is a library of handwritten digits.

Prediction challenges

The MNIST dataset is a library of handwritten digits.

In a prediction challenge, you are given a training set of images of handwritten digits, which are labeled from 0 to 9.

Prediction challenges

The MNIST dataset is a library of handwritten digits.

In a prediction challenge, you are given a training set of images of handwritten digits, which are labeled from 0 to 9.

You are also given a test set of handwritten digits, which are not identified.

Prediction challenges

The MNIST dataset is a library of handwritten digits.

In a prediction challenge, you are given a training set of images of handwritten digits, which are labeled from 0 to 9.

You are also given a test set of handwritten digits, which are not identified.

Your job is to assign a digit to each image in the test set.

The Netflix prize

Netflix popularized prediction challenges by organizing an open,
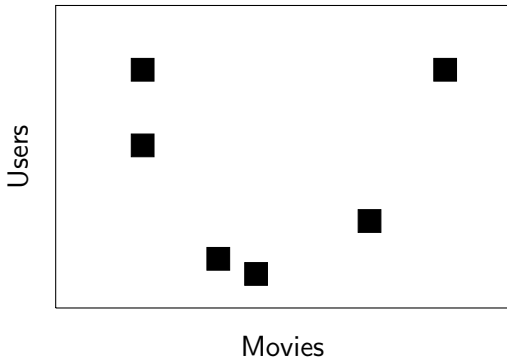blind contest to improve its recommendation system.

The prize was \$1 million.

The Netflix prize

Netflix popularized prediction challenges by organizing an open, blind contest to improve its recommendation system.

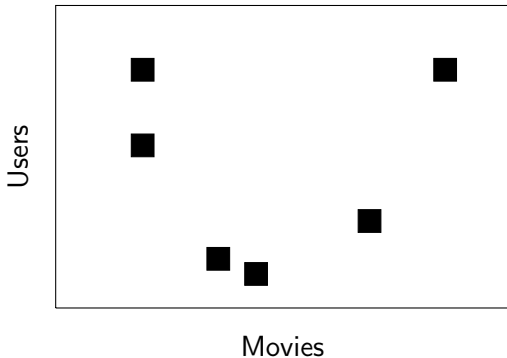**The prize was \$1 million.**

Some rankings were hidden in the training data

The Netflix prize

Netflix popularized prediction challenges by organizing an open, blind contest to improve its recommendation system.

The prize was $1 million.

The challenge was to predict those rankings

Kaggle

Company founded in 2010.

Business model:
- ▶ Organize prediction competitions hosted online.
- ▶ Offer companies consulting services from Kaggle "stars".

Kaggle

Company founded in 2010.

Business model:
- ▶ Organize prediction competitions hosted online.
- ▶ Offer companies consulting services from Kaggle "stars".

Kaggle-in-class is a competition engine offered to degree-granting institutions for free. Stats 202 was the first class to use it!