# Lecture 1

## Machine Learning Basics

Sung Kim <hunkim+ml@gmail.com>

# Basic concepts

- What is ML?
- What is learning?
  - supervised
  - unsupervised
- What is regression?
- What is classification?

# Machine Learning

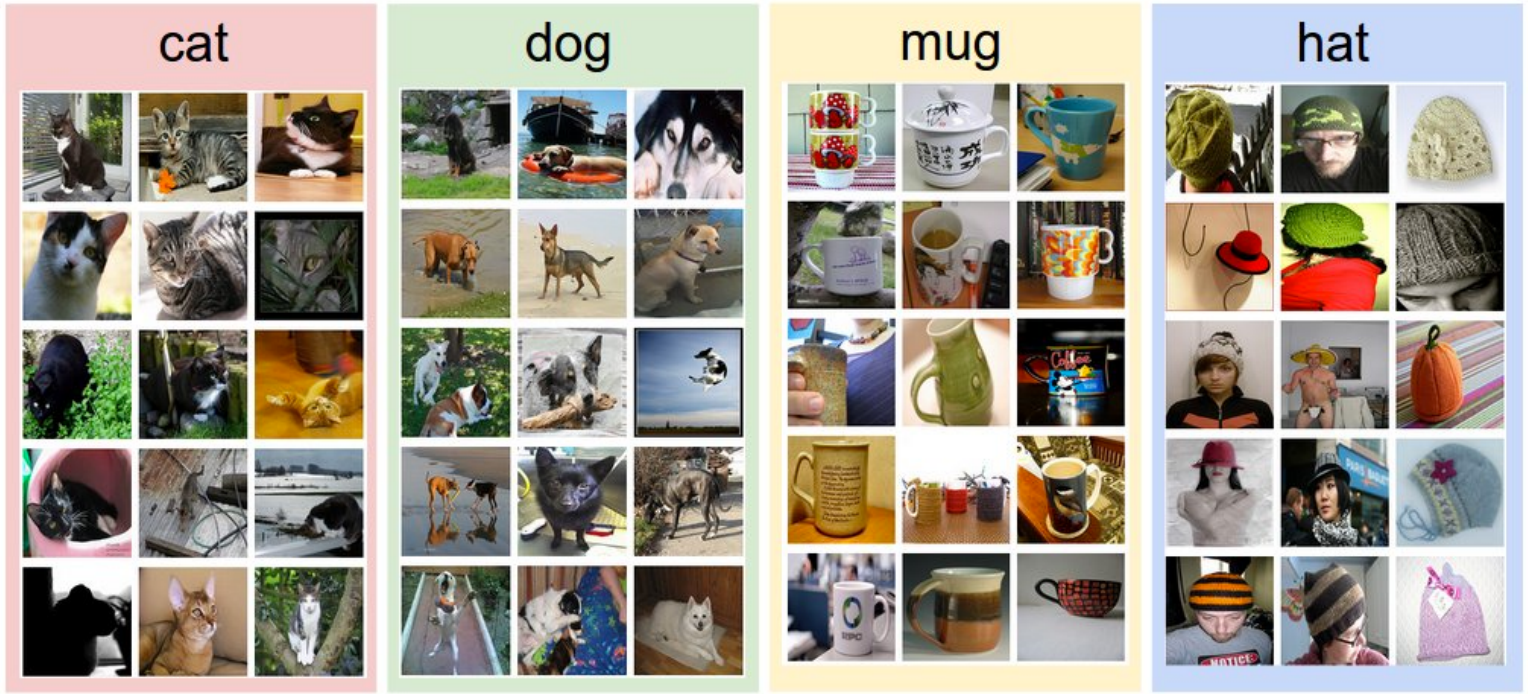
- Limitations of explicit programming
  - Spam filter: many rules
  - Automatic driving: too many rules
- Machine learning: "Field of study that gives computers the ability to learn without being explicitly programmed" Arthur Samuel (1959)

# Supervised/Unsupervised learning

- Supervised learning:
  - learning with labeled examples - training set

# Supervised learning

An example training set for four visual categories.




http://cs231n.github.io/classification/

# Supervised/Unsupervised learning

- Supervised learning:
  - learning with labeled examples
- Unsupervised learning: un-labeled data
  - Google news grouping
  - Word clustering

# Supervised learning

- Most common problem type in ML
  - **Image labeling:** learning from tagged images
  - **Email spam filter**: learning from labeled (spam or ham) email
  - **Predicting exam score**: learning from previous exam score and time spent

# Training data set

# AlphaGo

# Types of supervised learning

- Predicting final exam score based on time spent
  - regression
- Pass/non-pass based on time spent
  - binary classification
- Letter grade (A, B, C, E and F) based on time spent
  - multi-label classification

# Predicting final exam score based on time spent

| x (hours) | y (score) |
|---|---|
| 10 | 90 |
| 9 | 80 |
| 3 | 50 |
| 2 | 30 |

# Pass/non-pass based on time spent

| x (hours) | y (pass/fail) |
|---|---|
| 10 | P |
| 9 | P |
| 3 | F |
| 2 | F |

# Letter grade (A, B, …) based on time spent

| x (hours) | y (grade) |
|---|---|
| 10 | A |
| 9 | B |
| 3 | D |
| 2 | F |

**Next**

Linear regression