Machine Learning Project

What is a project?

Final project

- Novel implementation and evaluation of a machine learning algorithm

- Apply ML to a problem that interests you, using a publicly available dataset

- Milestones (see schedule for estimated deadlines)
  - Proposal
  - Two updates
  - Final presentation (in class)*
  - Final write up (conference paper style)

*demos are welcome but not necessary; quantitative evaluation much more important

Projects should use existing dataset

- Find a dataset online, implement and apply your own algorithm
  - See class website for resources

- Apply algorithms learned in class on to Kaggle competition
  - http://www.kaggle.com/competitions
  - Multiple teams can compete
- Projects proposing data collection will not be accepted

GALAXY ZOO

2 months
176 teams
$16,000

DENIED
APPROVED

46 days
166 teams
$10,000

Teams

- Teams of 3 (self-organize, or we can assign)
- Document what each project team member is responsible for, e.g.,
  - Dataset
  - Background research
  - Final write up
  - Lines of code written
  - Sections of paper/presentation
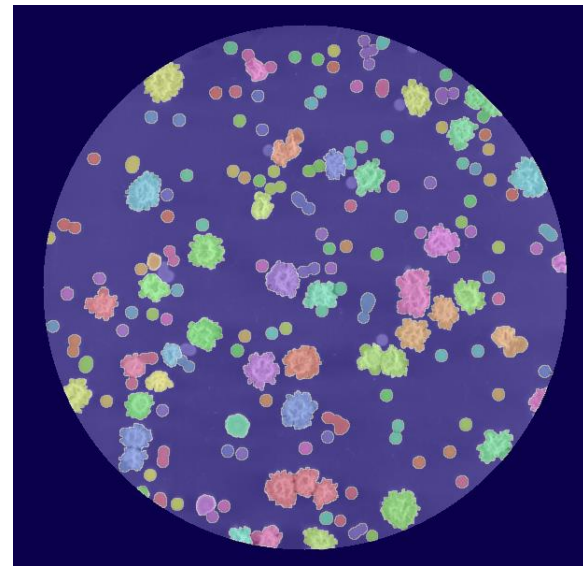 - Several of the above, etc...
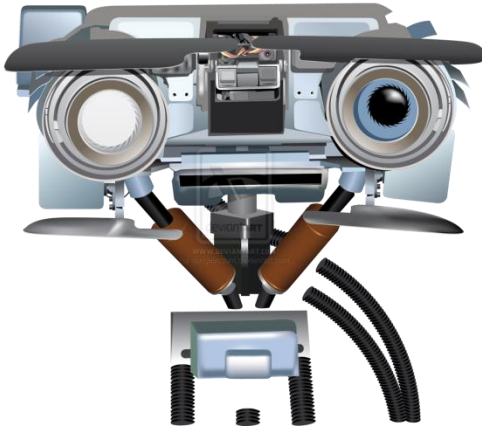
Project Examples
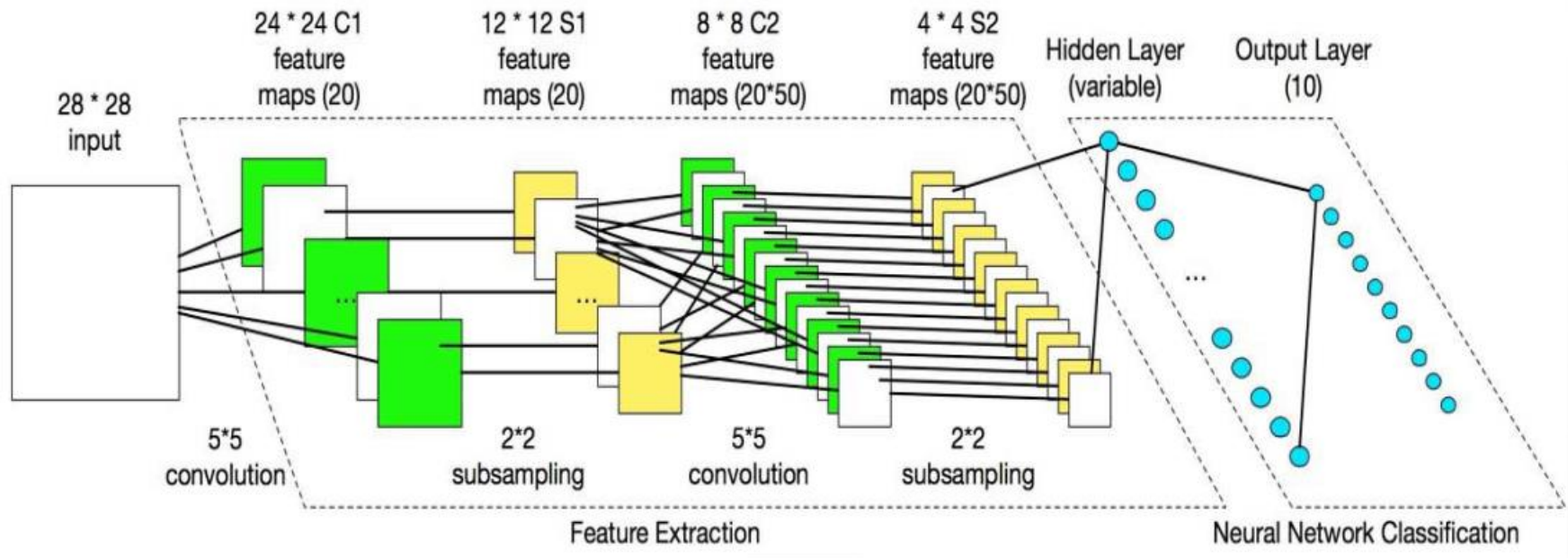
Examples from past courses

Bacteria Colony Morphology Classifier

- Many harmful bacteria form communities known as biofilms

- Biofilms are highly resistant to antibiotics and therefore pose a serious medical threat

- Goal: Classify bacterial colonies based on their respective biofilm morphologies
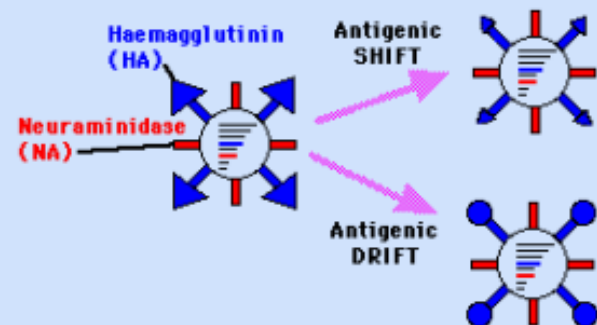
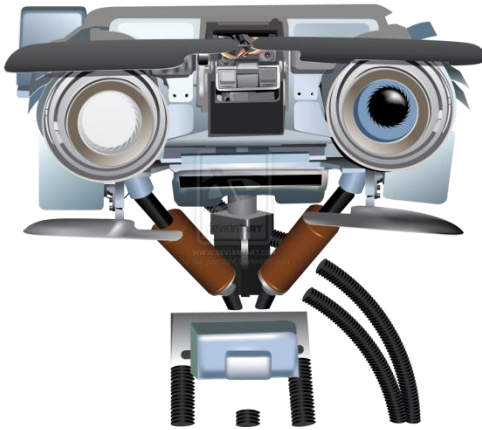# Neural Network for Digit Classification

Predicting H1N1 virus transmissibility

- Influenza is an infectious disease of birds and mammals caused by the virus of Orthomyxoviridae family.

- Avian ⟹ Human

- Hemagglutinin (HA) sequence
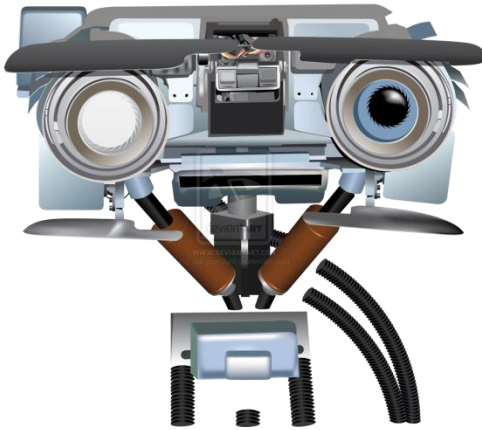
Machine Learning Project

What is a proposal?

Project Proposal

- List of team members/roles
- Two page description:
  - Goal: what is the problem?
  - Motivation: why should we care about the problem? Include a literature survey
  - Approach: a rough initial idea
  - Dataset: detailed dataset description
  - Expected outcome

What can I do now?

- Talk to perspective team members
  - Use Google group!

- Decide on (several) dataset(s)

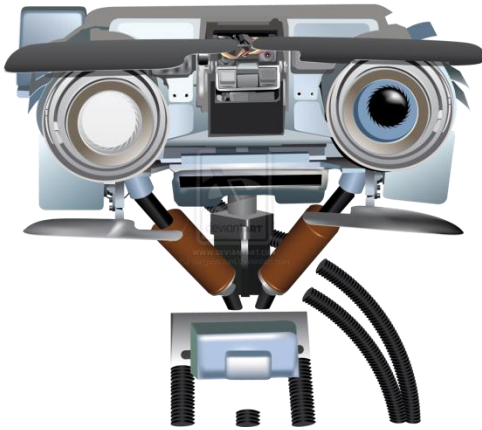- Start reading research papers that have tackled this problem

Machine Learning Project

What is a presentation?

Final Presentation

- 5 minutes per team

- I will determine team order

- Focus on
  - explaining the main problem and solution clearly,
  - not on providing lots of detail
  - do show results/evaluation

Machine Learning Project

What is final report?

Components of a final report

- Title and author(s) follow provided formatting template
- Goal   first paragraph of the introduction
  - what problem does your method solve? e.g. classify tumors as malignant or benign.
- Introduction   ~1 page
  - why should we care about the **problem**?
  - why should we care about the **method**?
- Background   ~ ½ page
  - what have others already tried? Cite related work and explain.
- Approach   ~2 pages
  - what is your method? Explain all variables and formulas
- Dataset   ~½ page
  - what data was used? Include examples if possible
- Evaluation   ~3 pages
  - feature extraction, experiment setup and results
- Conclusion   ~ paragraph
  - what is the take-away message?
- Team roles ~ paragraph
- References   ~ page

These page guidelines assume a total of about 8 pages. They are just guidelines, not strict requirements.