



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Artificial Intelligence

Lecturer 13 – Machine Learning

School of Information and Communication
Technology - HUST

Introduction of Machine learning

- Definitions of Machine learning...
 - A process by which a system improves its performance [Simon, 1983]
 - Any computer program that improves its performance at some task through experience [Mitchell, 1997]
 - Programming computers to optimize a performance criterion using example data or past experience [Alpaydin, 2004]
- Representation of the learning problem [Mitchell, 1997]

Learning = Improving with experience at some task

 - Improve over task **T**
 - With respect to performance measure **P**
 - Based on experience **E**

Application examples of ML (1)

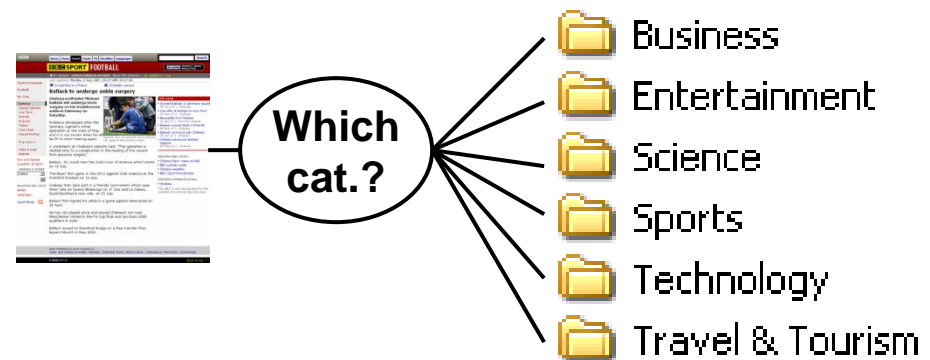
Web pages filtering problem

- **T**: to predict which Web pages a given user is interested in
- **P**: % of Web pages correctly predicted
- **E**: a set of Web pages identified as interested/uninterested for the user



Web pages categorization problem

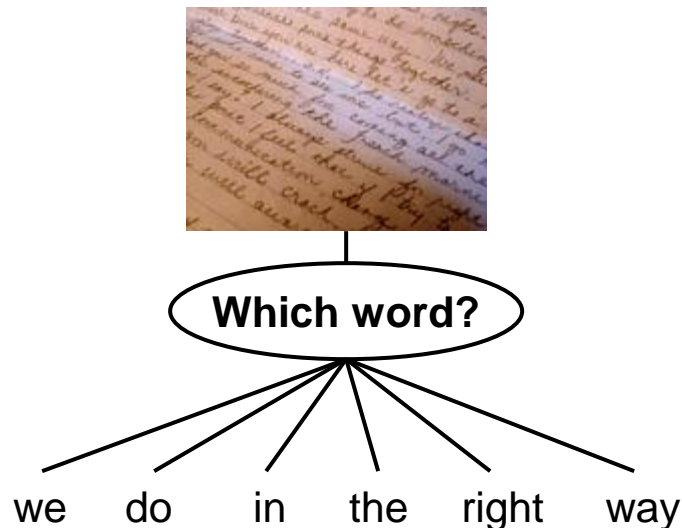
- **T**: to categorize Web pages in predefined categories
- **P**: % of Web pages correctly categorized
- **E**: a set of Web pages with specified categories



Application examples of ML (2)

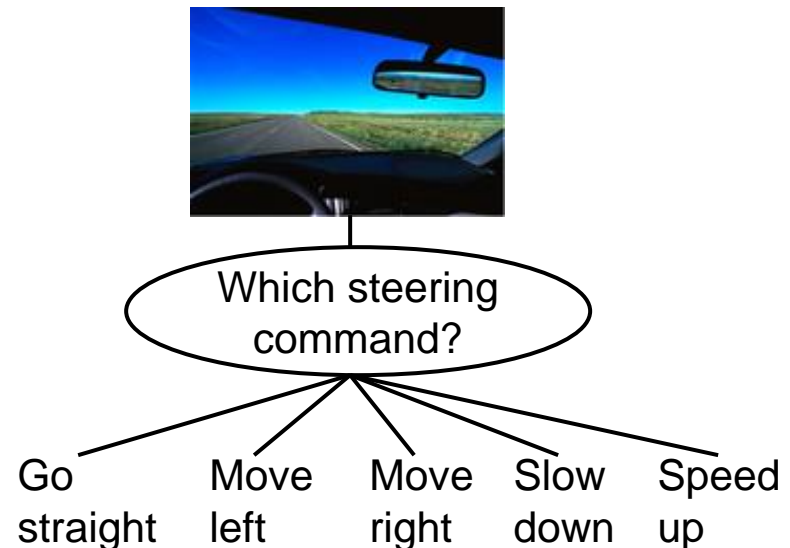
Handwriting recognition problem

- **T**: to recognize and classify handwritten words within images
- **P**: % of words correctly classified
- **E**: a database of handwritten words with given classifications (i.e., labels)



Robot driving problem

- **T**: to drive on public highways using vision sensors
- **P**: average distance traveled before an error (as judged by human overseer)
- **E**: a sequence of images and steering commands recorded while observing a human driver



Key elements of a ML problem (1)

- Selection of the training examples
 - Direct or indirect training feedback
 - With teacher (i.e., with labels) or without
 - The training examples set should be representative of the future test examples
- Choosing the target function (a.k.a. hypothesis, concept, etc.)
 - $F: X \rightarrow \{0,1\}$
 - $F: X \rightarrow$ a set of labels
 - $F: X \rightarrow \mathbb{R}^+$ (i.e., the positive real numbers domain)
 - ...

Key elements of a ML problem (2)

- Choosing a representation of the target function
 - A polynomial function
 - A set of rules
 - A decision tree
 - A neural network
 - ...
- Choosing a learning algorithm that learns (approximately) the target function
 - Regression-based
 - Rule induction
 - ID3 or C4.5
 - Back-propagation
 - ...

Issues in Machine Learning (1)

- Learning algorithm
 - Which algorithms can approximate the target function?
 - Under which conditions does a selected algorithm converge (approximately) to the target function?
 - For a certain problem domain and given a representation of examples which algorithm performs best?
- Training examples
 - How many training examples are sufficient?
 - How does the size of the training set influence the accuracy of the learned target function?
 - How does noise and/or missing-value data influence the accuracy?

Issues in Machine Learning (2)

- Learning process
 - What is the best strategy for selecting a next training example? How do selection strategies alter the complexity of the learning problem?
 - How can prior knowledge (held by the system) help?
- Learning capability
 - What target function should the system learn?

Representation of the target function: expressiveness vs. complexity
 - What are the theoretical limits of learnability?
 - How can the system generalize from the training examples?

To avoid the overfitting problem
 - How can the system automatically alter its representation?

To improve its ability to represent and learn the target function

Types of learning problems

- A rough (and somewhat outdated) classification of learning problems:
 - Supervised learning, where we get a set of training inputs and outputs
 - classification, regression
 - Unsupervised learning, where we are interested in capturing inherent organization in the data
 - clustering, density estimation
 - Reinforcement learning, where we only get feedback in the form of how well we are doing (not what we should be doing)
 - Planning