Lecture 1 Introduction to ML

Zhihua Jiang

Reminder

- Please join the QQ group: 757038610
- Teacher's email: tjiangzhh@jnu.edu.cn / 22799204@qq.com

Grade System

1. Attendance
2. Homework including after-class exercises and a project

   Suggestions for doing the project:

   - September — problem definition, data collection, and feature extraction
   - October – model training, evaluations and experimental analysis
   - November/December – presentation and the final report

3. Final exam (closed-book)

Reference Materials

There is no required textbook for this course. The following books are recommended as optional reading:

- <The Elements of Statistical Learning: Data Mining, Inference, and Prediction> Trevor Hastie , World Book Inc, ISBN：9787510084508
- 《机器学习》周志华著，清华大学出版社，2016年01月
- Stanford CS229 – Machine Learning (lectured by **Andrew Ng**)

(https://see.stanford.edu/Course/CS229)

CS229 - Machine Learning

Course Details Show All

Course Description

This course provides a broad introduction to machine learning and statistical pattern recognition.

Topics include: supervised learning (generative/discriminative learning, parametric/non-parametric learning, neural networks, support vector machines); unsupervised learning (clustering, dimensionality reduction, kernel methods); learning theory (bias/variance tradeoffs; VC theory; large margins); reinforcement learning and adaptive control.
The course will also discuss recent applications of machine learning, such as to robotic control, data mining, autonomous navigation, bioinformatics, speech recognition, and text and web data processing.
Students are expected to have the following background:

Prerequisites: - Knowledge of basic computer science principles and skills, at a level sufficient to write a reasonably non-trivial computer program.
- Familiarity with the basic probability theory. (Stat 116 is sufficient but not necessary.)
- Familiarity with the basic linear algebra (any one of Math 51, Math 103, Math 113, or CS 205 would be much more than necessary.)

- Syllabus
- DOWNLOAD All Course Materials

Instructor +

Handouts +

Resources +

Assignments +

Course Sessions (20): Show All

Lecture 1 +

Lecture 2 +

Lecture 3 +

Lecture 4 +

Lecture 5 +

Lecture 6 +

Lecture 7 +

Lecture 8 +

Lecture 9 +

Lecture 10 +

Lecture 11 +

Lecture 12 +

Lecture 13 +

Lecture 14 +

Lecture 15 +

Lecture 16 +

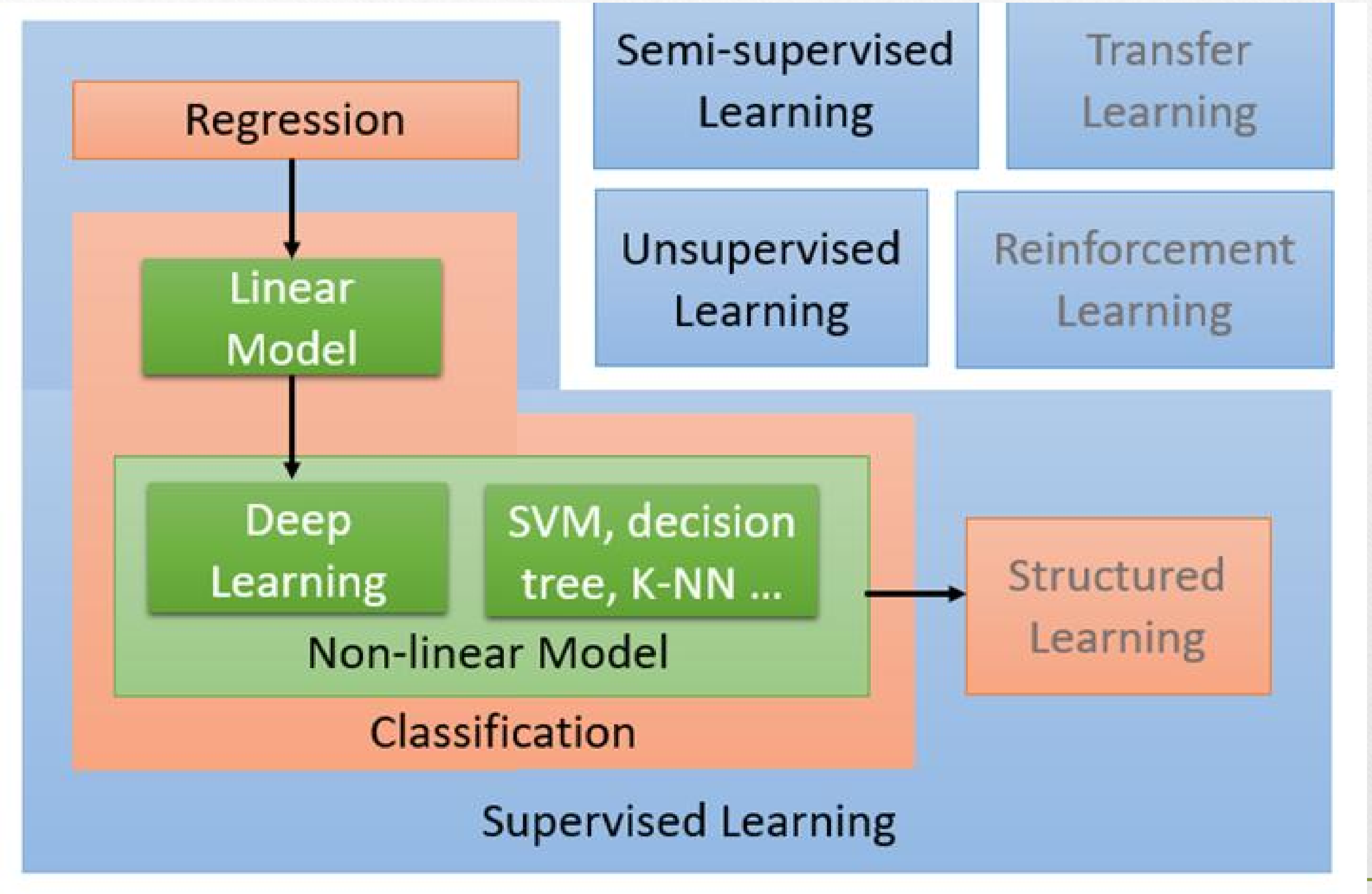
Lecture 17 +

Lecture 18 +

Lecture 19 +

Lecture 20 +

ML Map


Semi-supervised Learning
Transfer Learning
Regression
Linear Model
Unsupervised Learning
Reinforcement Learning
Deep Learning
SVM, decision tree, K-NN ...
Non-linear Model
Structured Learning
Classification
Supervised Learning

Outline

1. Introduction
2. Supervised Learning
3. Linear Model
4. Regression
5. Decision tree
6. Neural network (NN)
7. Support vector machine (SVM)
8. Ensemble learning
9. Semi-supervised learning & unsupervised learning

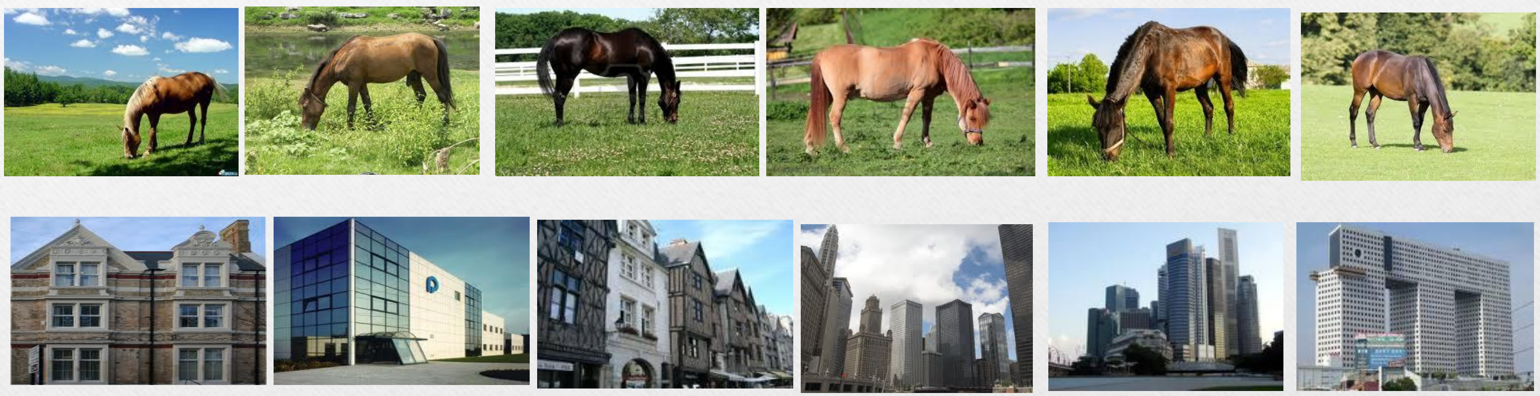
Introduction

1. Why ML
2. History of ML
3. Mathematical basis of ML
4. Basic terminology
5. Inductive bias

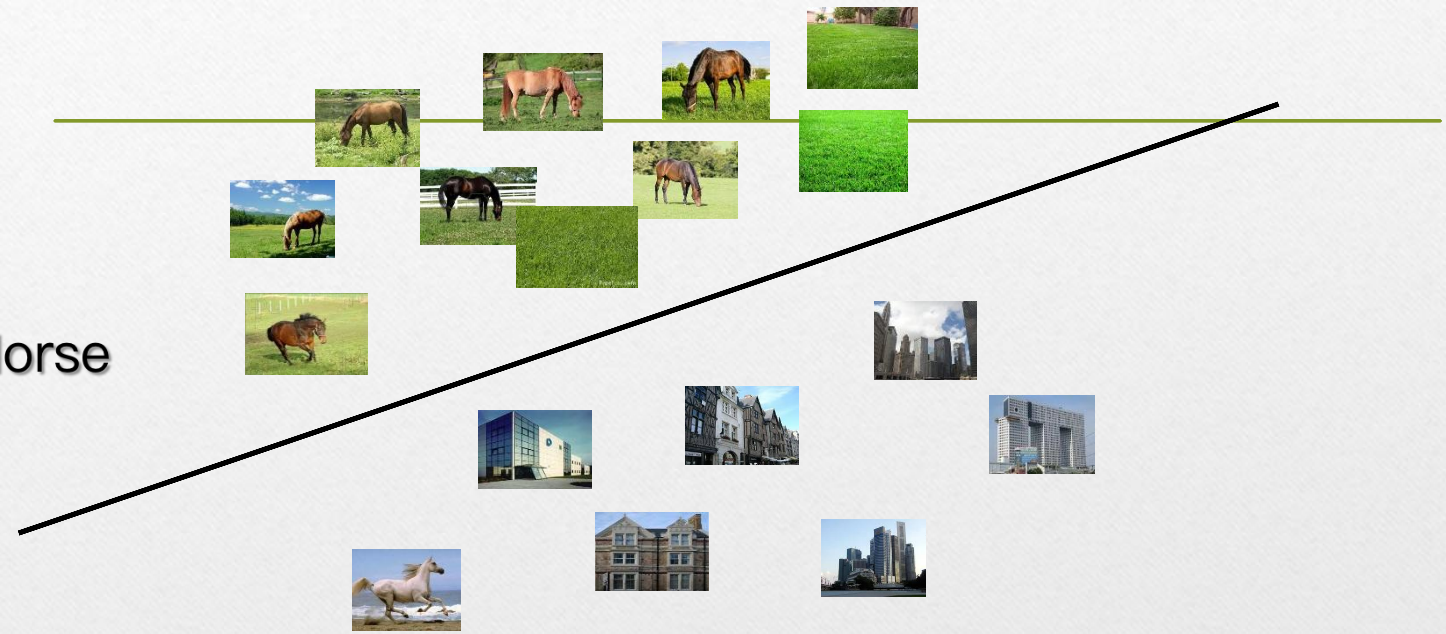
1. Why ML

- What is ML?
    - learn from data
- What can ML do?
    - Prediction
    - Clustering
    - Making decision
    - Knowledge discovery
    - ……

Example: Classification



Classifier



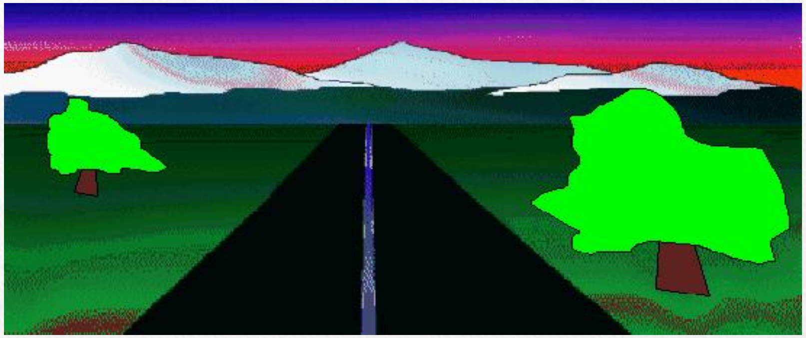
Horse

Non Horse

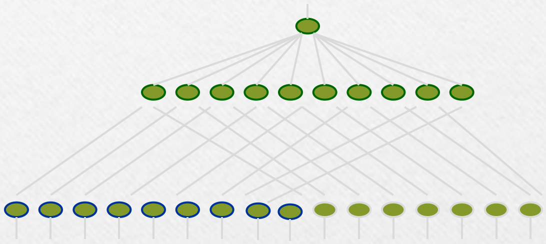
Evaluation metric: accuracy

Example: Auto Vehicle Navigation

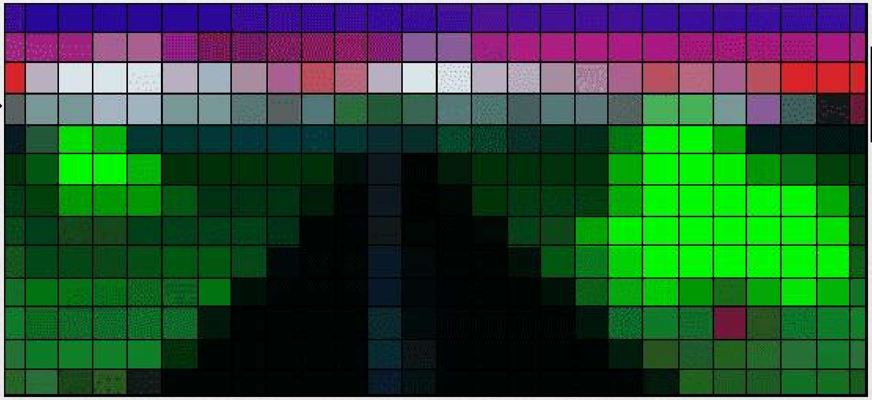
Raw data: image



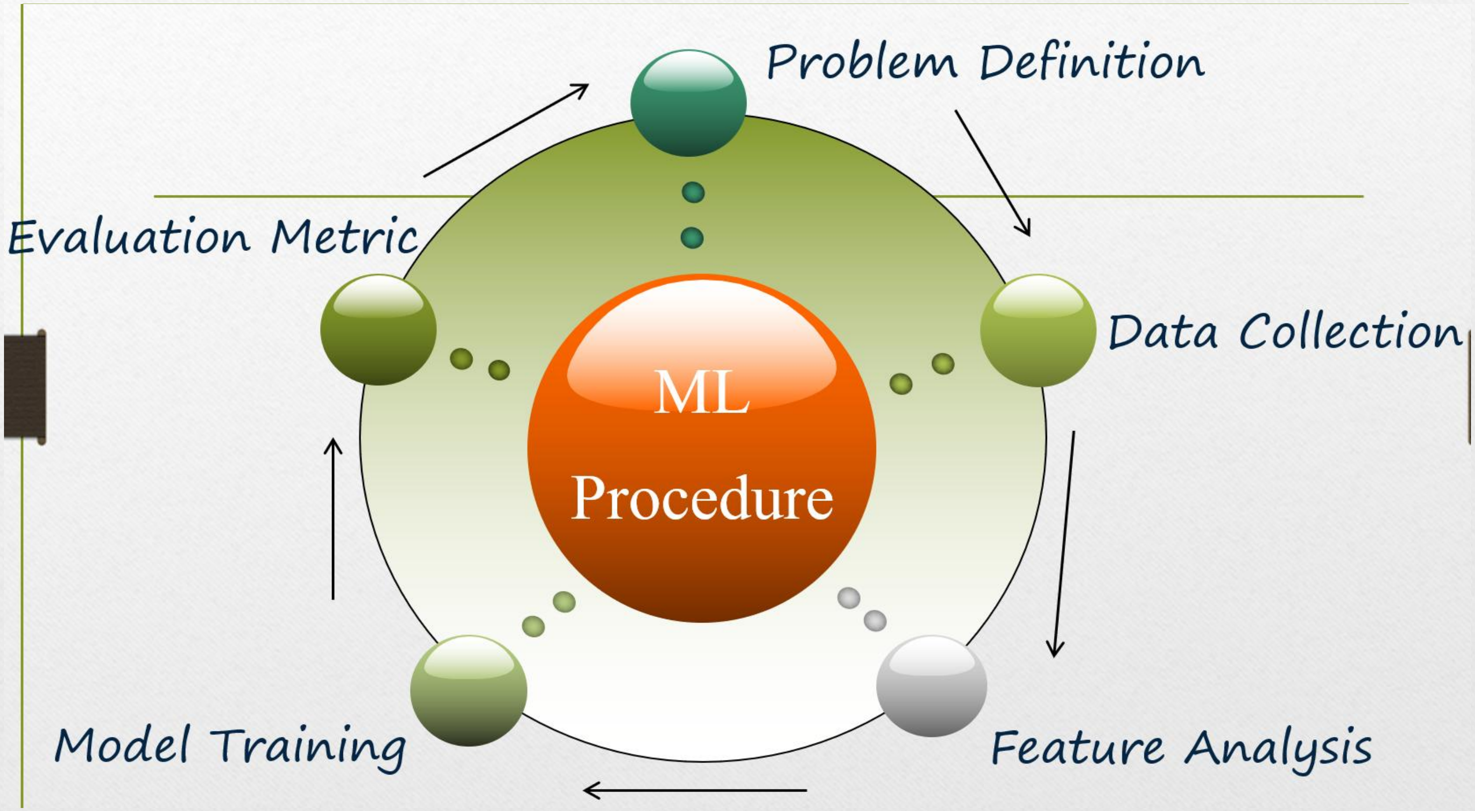
Output: steering direction



Input: pixel matrix



General Procedure in ML



2. History of ML

ML comes from AI (Artificial Intelligence, 1956–)

- 1st phase: (1956–1960s) logic reasoning

  e.g., Automatic theorem proving
- 2nd phase: (1970s–1980s) knowledge engineering

  e.g., Expert system
- 3rd phase: (1990s–now) machine learning

  e.g., Symbolism; Connectionism; Statistical learning

Symbolism

- Expert system
- Inductive logic programming
- Decision tree
- …

Connectionism

- Perceptron
- Neural network
- Deep learning
- …

Statistical learning

- Support vector machine
- Kernel methods
- …

3. Mathematical basis of ML

- Linear algebra
- Probability and statistics
- Differential calculus
- Algorithm and complexity

Linear algebra	Probability and statistics	Differential calculus	Algorithm and complexity
• principal component analysis • (主成分分析，PCA) • singular value decomposition (奇异值分解，SVD) • vector space and norms • (向量空间和范数)	• expectation and variance (期望和方差) • Bayes theorem • (贝叶斯定理) • maximum posteriori estimation • (最大后验估计)	• vector value function • (向量值函数) • gradient • (梯度)	• data structure • (数据结构) • stochastic algorithm • (随机算法) • duality method • (对偶方法)

Goal: optimization！

4. Basic terminology

- ESL Book (Sec. 2.2 Variable Types and Terminology)

We will typically denote an input variable by the symbol X. If X is a vector, its components can be accessed by subscripts X_j. Quantitative outputs will be denoted by Y, and qualitative outputs by G (for group). We use uppercase letters such as X, Y or G when referring to the generic aspects of a variable. Observed values are written in lowercase; hence the ith observed value of X is written as x_i (where x_i is again a scalar or vector). Matrices are represented by bold uppercase letters; for example, a set of N input p-vectors x_i, $i = 1, \ldots, N$ would be represented by the $N \times p$ matrix $\mathbf{X}$. In general, vectors will not be bold, except when they have N components; this convention distinguishes a p-vector of inputs x_i for the

4. Basic terminology

ith observation from the N-vector $\mathbf{x}_j$ consisting of all the observations on variable X_j. Since all vectors are assumed to be column vectors, the ith row of $\mathbf{X}$ is x_i^T, the vector transpose of x_i.

For the moment we can loosely state the learning task as follows: given the value of an input vector X, make a good prediction of the output Y, denoted by $\hat{Y}$ (pronounced "y-hat"). If Y takes values in $\mathbb{R}$ then so should $\hat{Y}$; likewise for categorical outputs, $\hat{G}$ should take values in the same set $\mathcal{G}$ associated with G.

4. Basic terminology

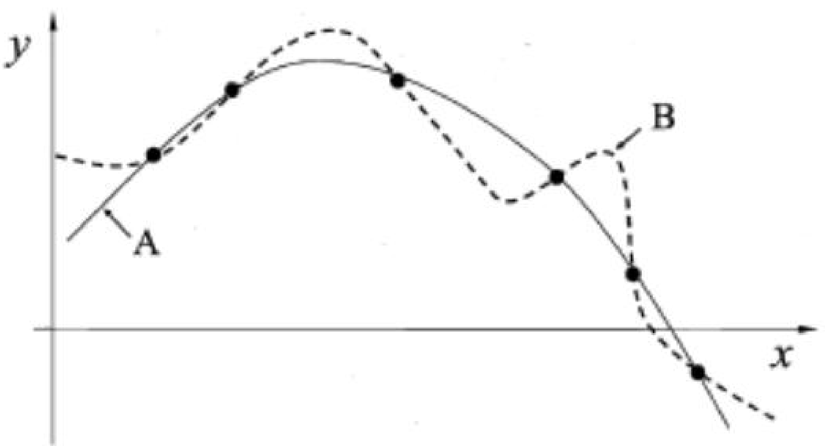
- Supervised learning: learn with labeled training data
- Unsupervised learning: learn with unlabeled training data
- Semi-supervised: a small amount of labeled data with a large amount of unlabeled data.
  - Train model with labeled data
  - Use the learned model to predict unlabeled data, then adjust parameter

5. Inductive bias

- Inductive bias (归纳偏好): any learning algorithm will prefer to some kind of hypothesis.

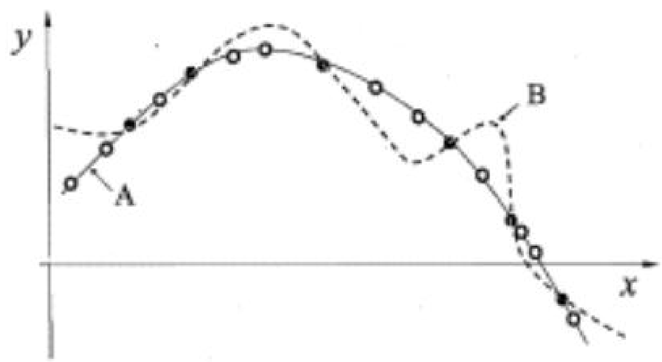
- Occam's razor (奥卡姆剃刀): the simplest is the best!

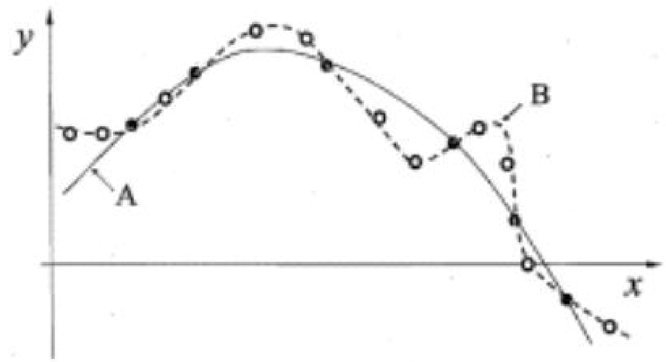
Ex. of Inductive bias



Q: If both A and B can fit the training set, which one is the best?

Q: Which one is better?





a) A is better than B

b) B is better than A

Hint: Black/White stands for training/testing example