Introduction to Machine Learning

Lecture 02: Supervised Learning I – Linear Regression

Sep. 19, 2024

Jie Wang

Machine Intelligence Research and Applications Lab

Department of Electronic Engineering and Information Science (EEIS)

http://staff.ustc.edu.cn/~jwangx/

jiewangx@ustc.edu.cn

Contents

- An Example: The ETA Problem
- Linear Regression by Least Squares
- Linear Regression by Maximum Likelihood

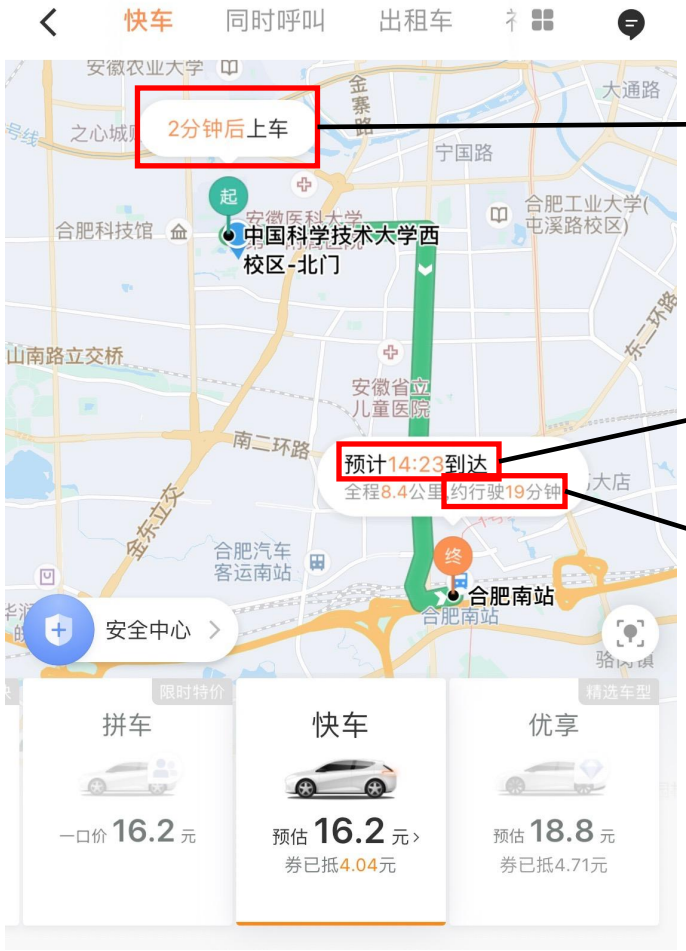
Some materials are from ESL, PRML, and RG.

- **An Example: The ETA Problem**

The ETA Problem

The **ETA** (**E**stimated **T**ime of **A**rrival) problem:

- Suppose that you are an engineer working at DiDi/Gaode/…. Your supervisor ask you to develop an algorithm to estimate the time of arrival for each customer. For the good of the customers' experience, the ETA given by your algorithm should be as accurate as possible.



Time of Wait

ETA

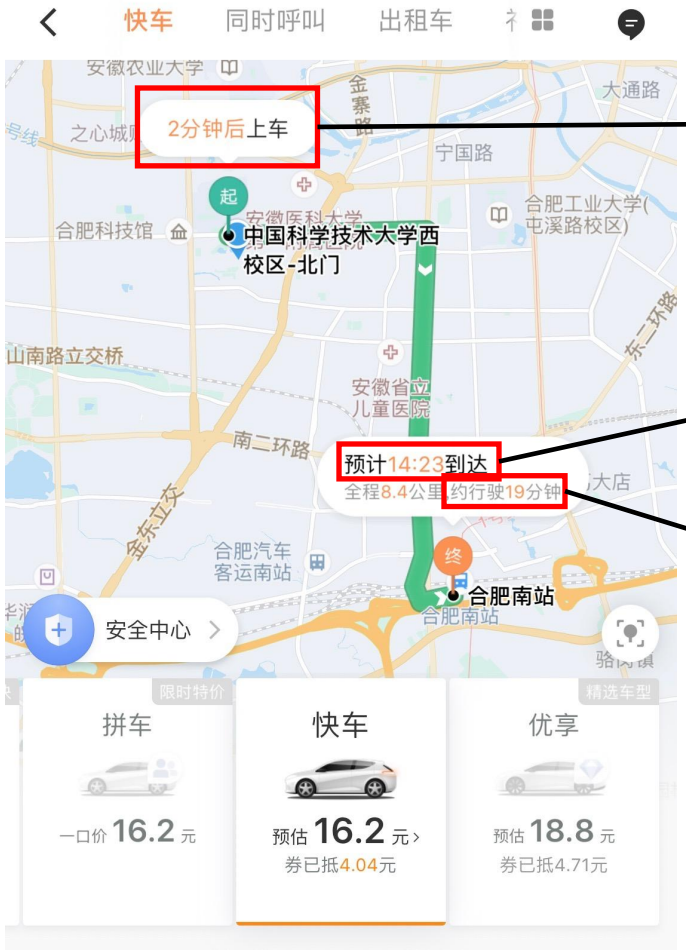
Time on the Road

ETA = Time of Wait + Time on the Road

The ETA Problem

The **ETA** (**E**stimated **T**ime of **A**rrival) problem:

- Suppose that you are an engineer working at DiDi/Gaode/…. Your supervisor ask you to develop an algorithm to estimate the time of arrival for each customer. For the good of the customers' experience, the ETA given by your algorithm should be as accurate as possible.



Time of Wait

ETA

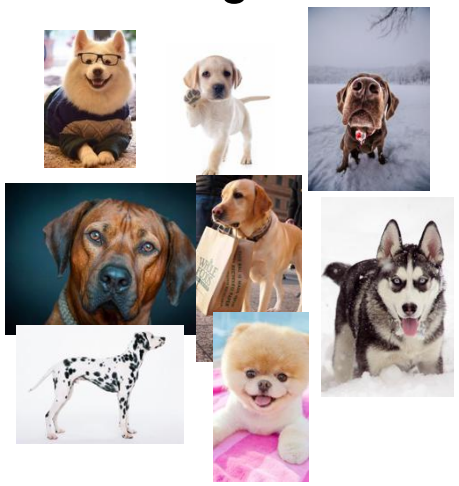
Time on the Road

ETA = Time of Wait + Time on the Road

Supervised Learning (Recall from Lec00)

labels

dog



cat

Training（Learning）

Each sample consists of an input x and a target output y.

Supervised Signal

Predicting

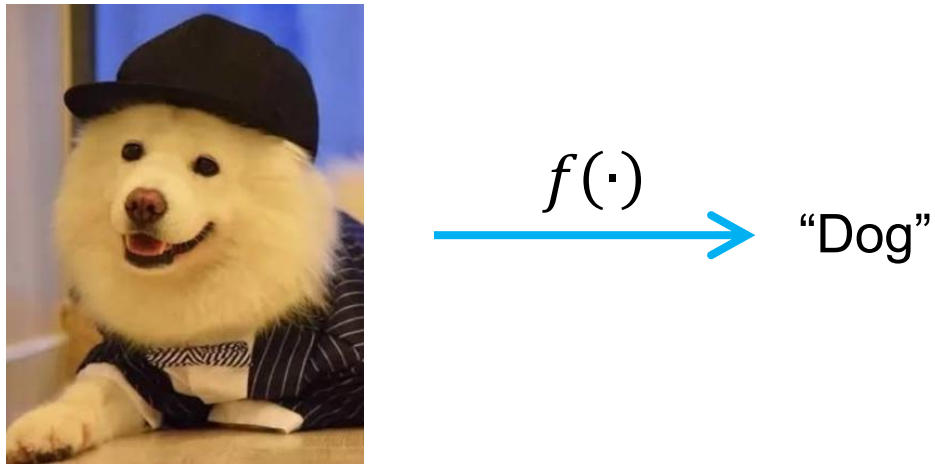
dog

- Regression : The output is a real number (vector)
- Classification : The output is a class label.

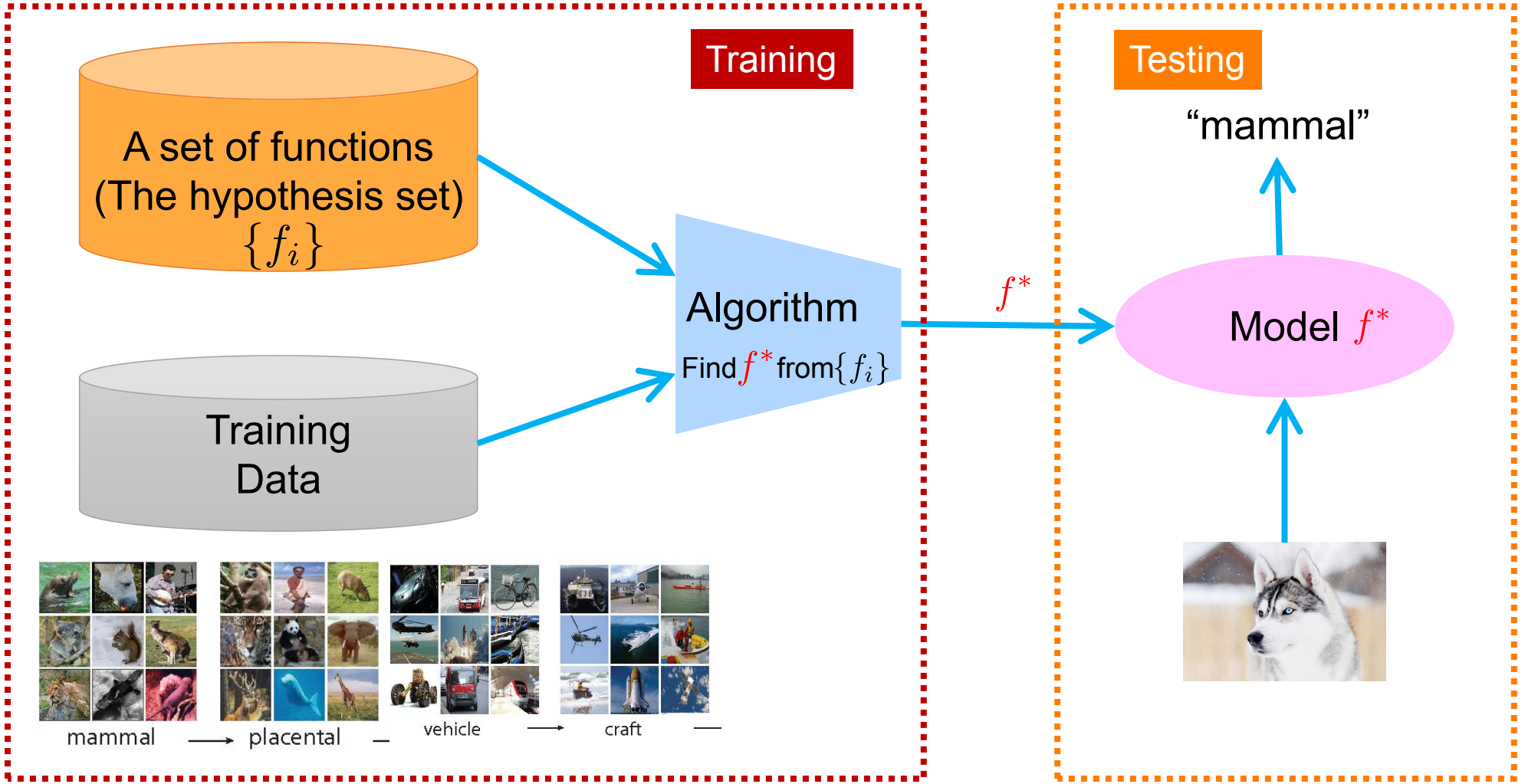
Supervised Learning (Recall from Lec00)

We are indeed looking for a mapping (function).

- Image classification

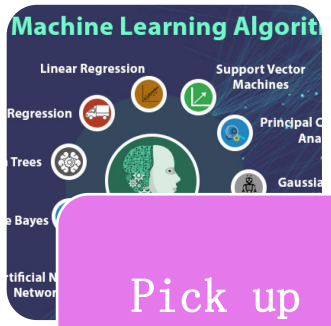


Framework of Supervised Learning (Recall from Lec00)

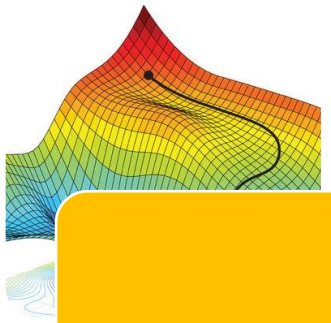


Framework of Supervised Learning

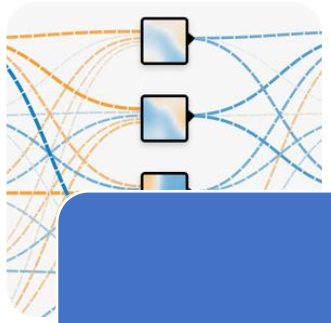
Data Preparation



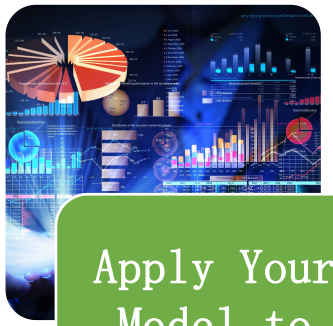
Pick up Hypothesis Set



Training



The Model



Apply Your Model to the Unseen Data

Step 1: Data Preparation

What kind of data you would like to collect?

Step 1: Data Preparation

- What kind of data you would like to collect?

	time	distance	origin	destination	…
Customer 1					
Customer 2					
Customer 3					
……					

Step 1: Data Preparation

- What kind of data you would like to collect?

Features

	time	distance	origin	destination	…
Customer 1					
Customer 2					
Customer 3					
……					

The data determines the upper bound of the performance that can be achieve by your model.

In real world applications, you need to determine which kind of data can be helpful to your task and collect them by yourself.

The knowledge that can help you to determine which kind of data to collect is the so-called ***domain knowledge***.

Step 1: Data Preparation

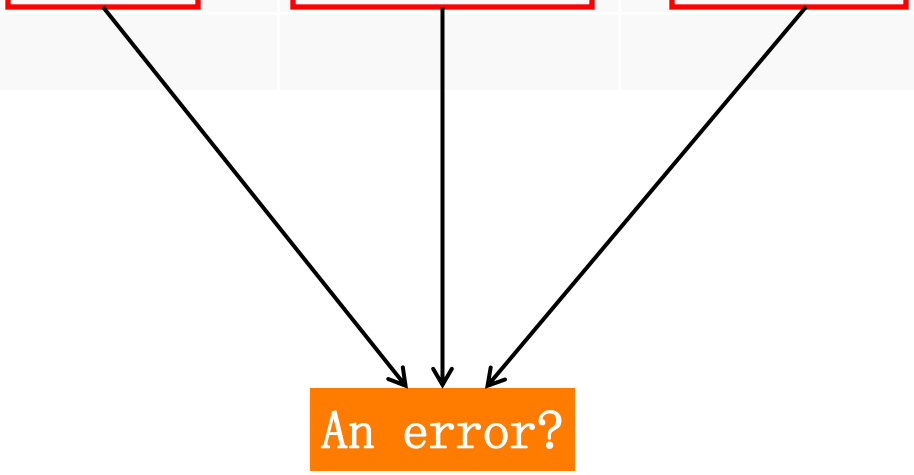
- Data cleaning

	time	distance	origin	destination	…
Customer 1					
Customer 2		?			
Customer 3				?	
……					

Step 1: Data Preparation

- Data cleaning

	time	distance	origin	destination	…
Customer 1					
Customer 2					
Customer 3		800 km	中科大西校区	合肥南站	
……					

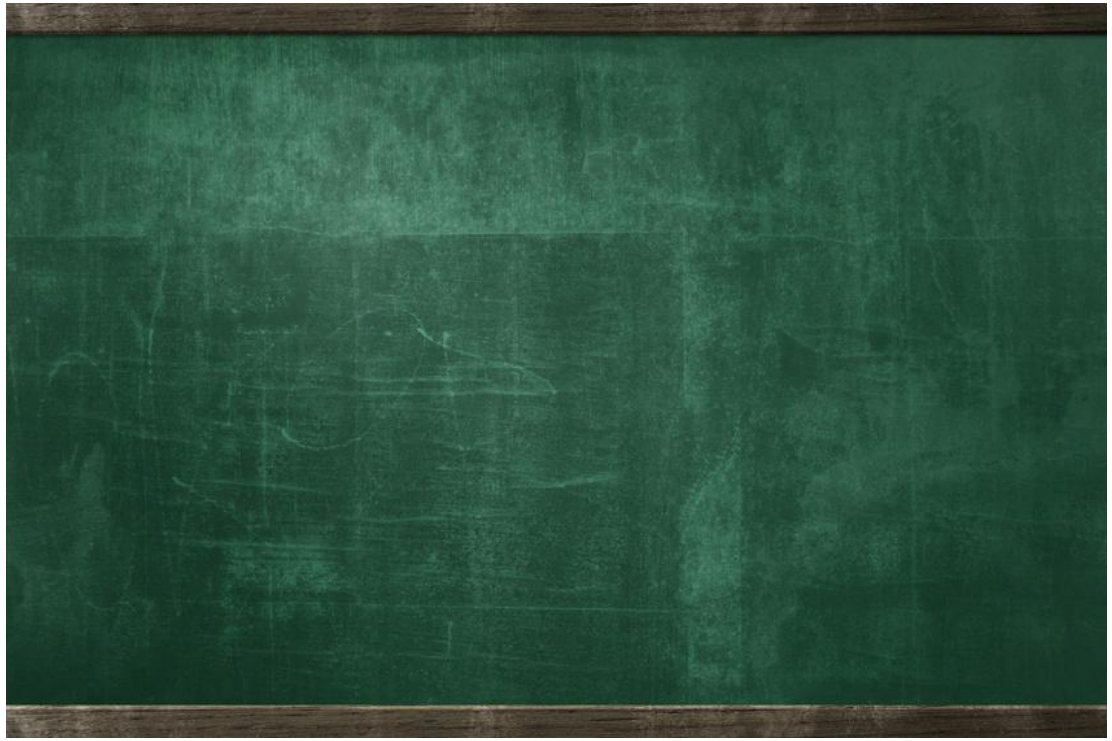


Step 1: Data Preparation

In real world applications, data preparation (cleaning) often takes up to **80%** (even **90%**) of the entire project lifecycle.

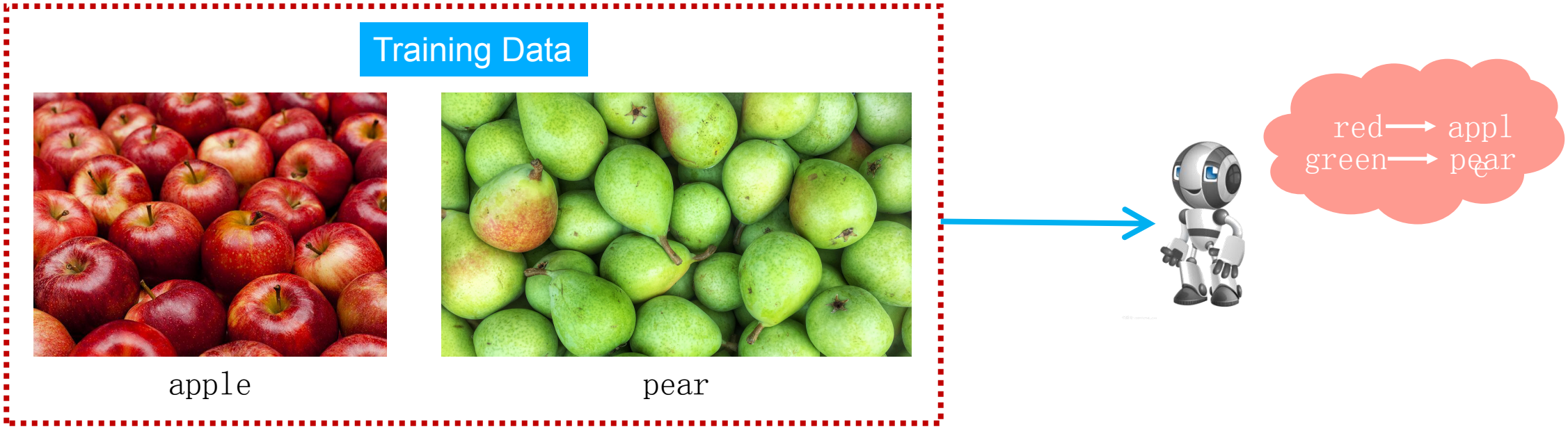
Step 2: Picking a hypothesis space

- Linear regression
  - Least squares



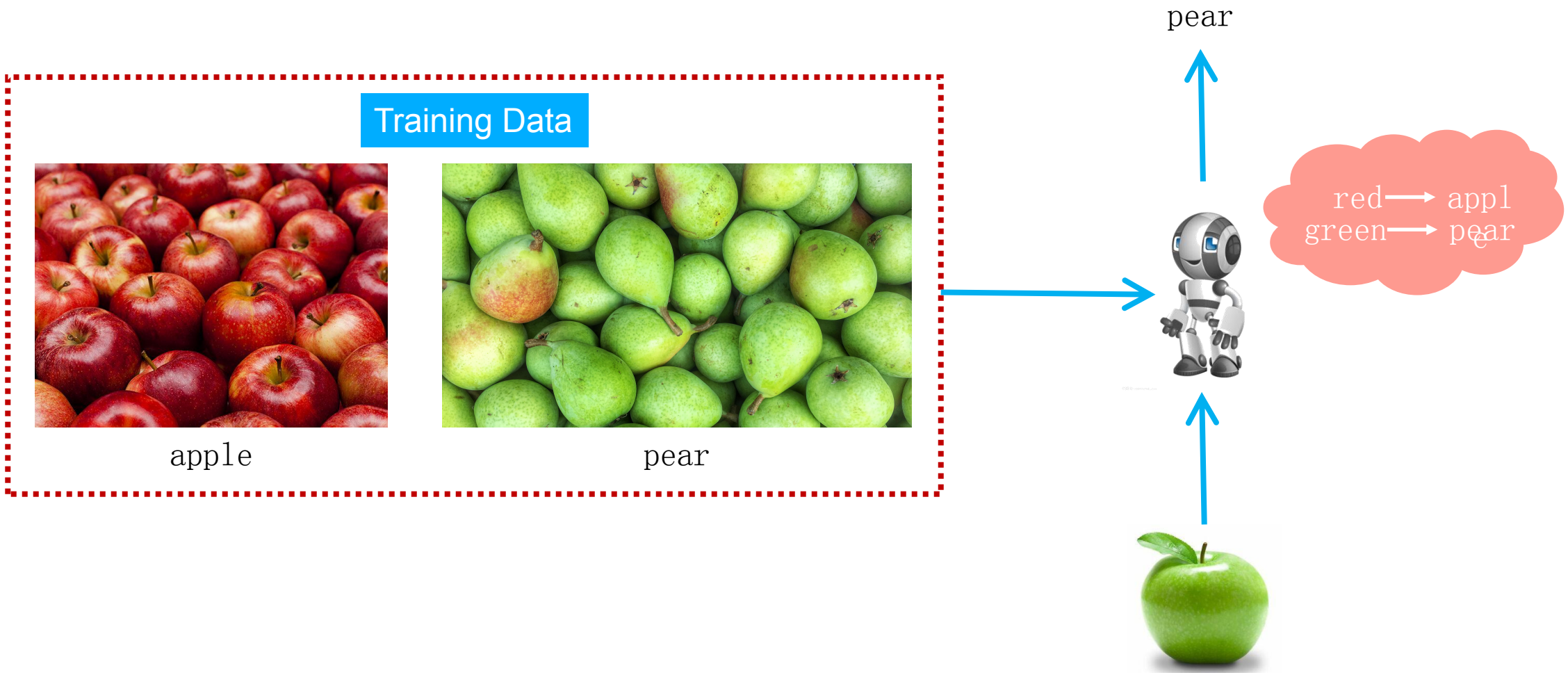
Step 2: Picking a hypothesis space

- Overfitting



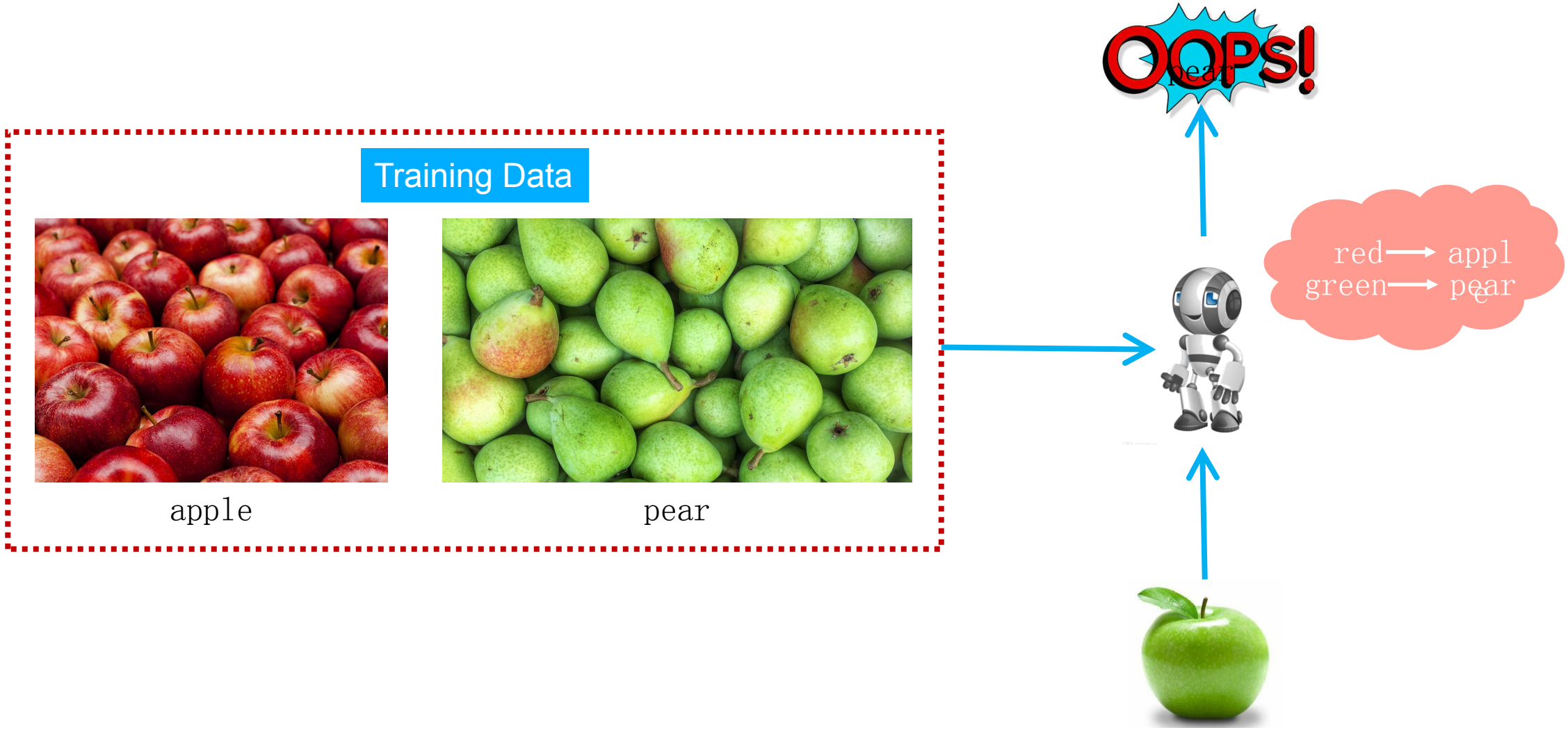
Step 2: Picking a hypothesis space

- Overfitting



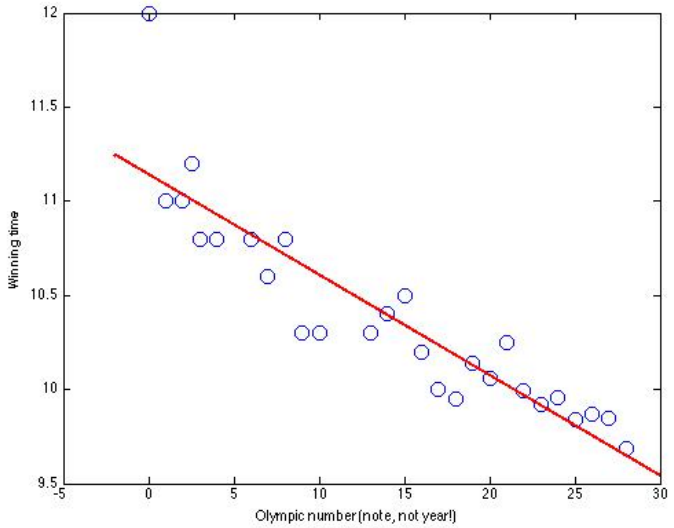
Step 2: Picking a hypothesis space

- Overfitting

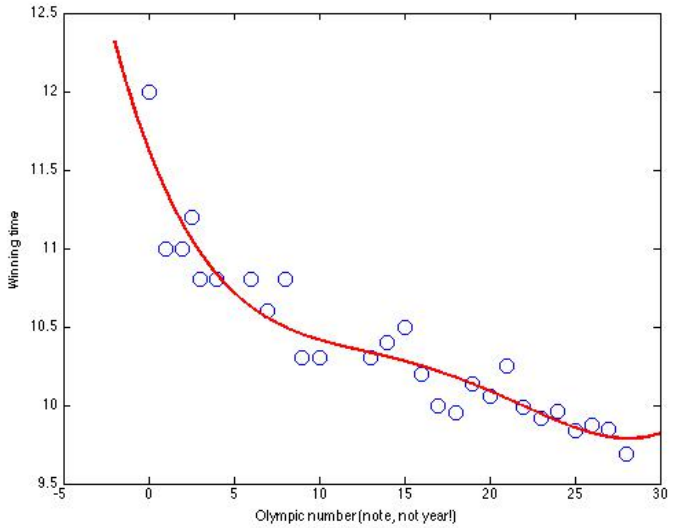


Step 2: Picking a hypothesis space

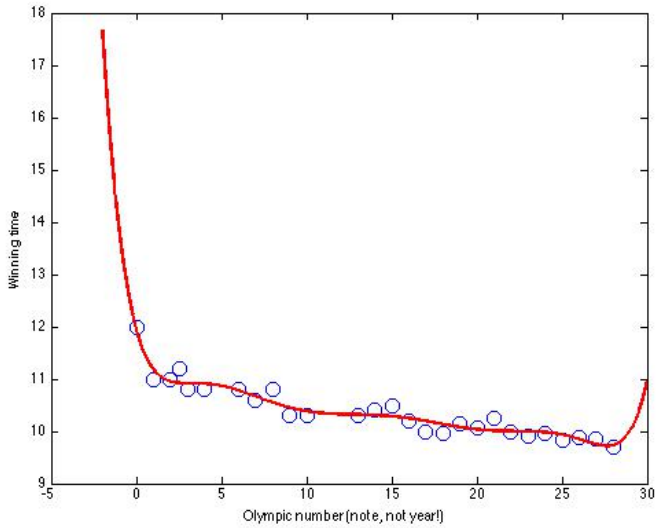
- Overfitting



Linear function



4th order polynomial

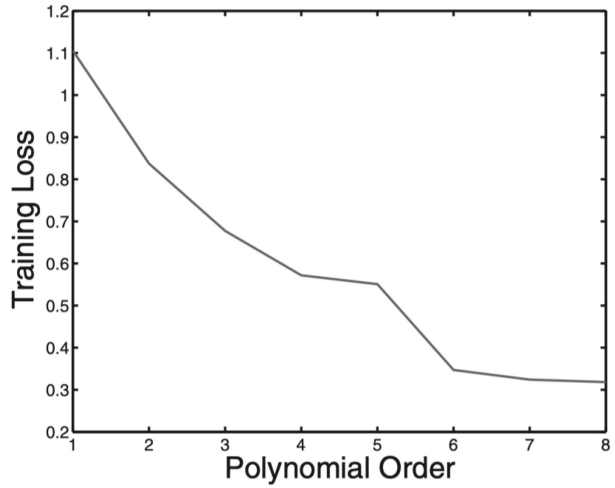


8th order polynomial

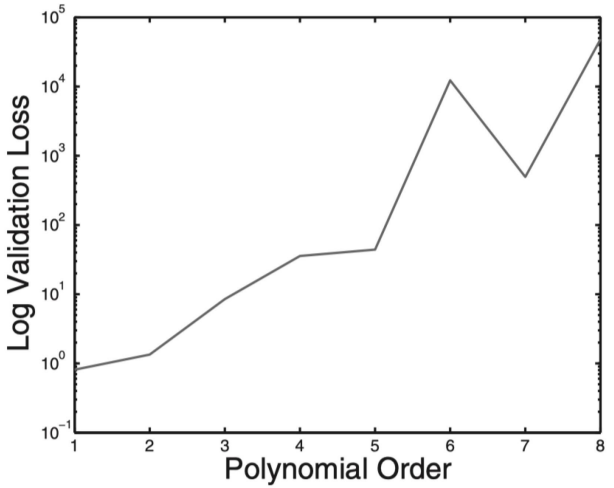
Step 2: Picking a hypothesis space

- How to alleviate overfitting?

Validation data Either provided separately or can be created by splitting the original data



(a) Training loss for the Olympic men's 100 m data.

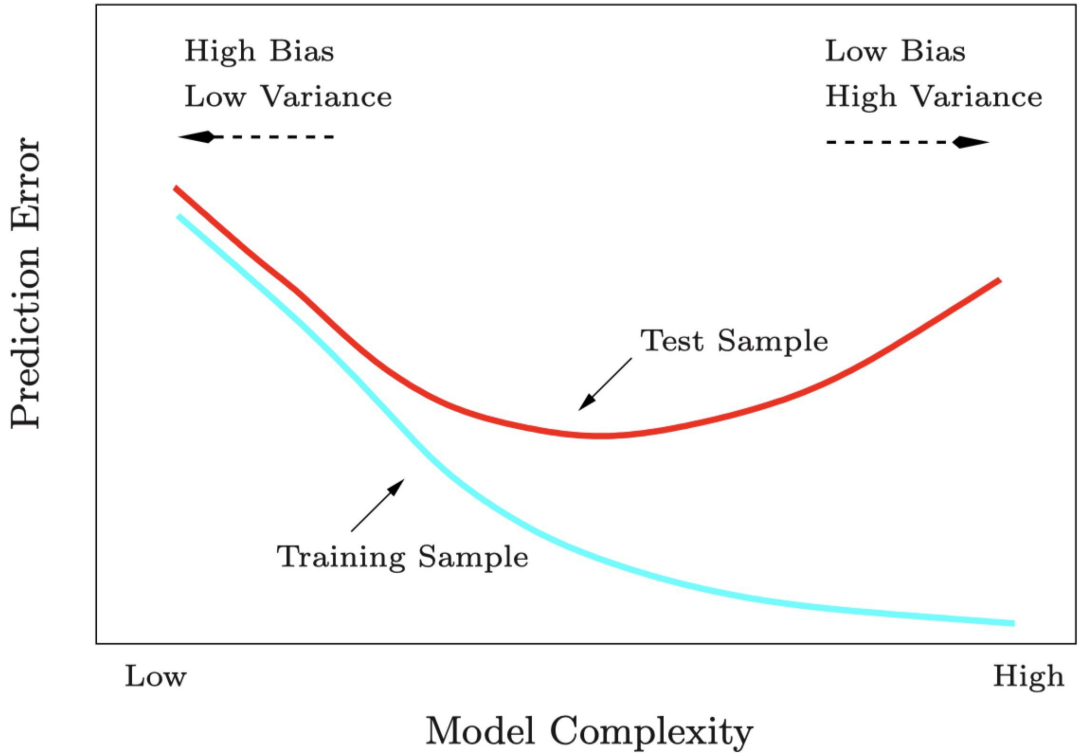


(b) Log validation loss for the Olympic men's 100 m data. When using the squared loss, this is also known as the squared predictive error and measures how close the predicted values are to the true values. Note that the log loss is plotted as the value increases so rapidly.

Step 2: Picking a hypothesis space

- How to alleviate overfitting?

Validation data — Either provided separately or can be created by splitting the original data



Step 2: Picking a hypothesis space

- How to alleviate overfitting?

Validation data → Cross-validation

validation	train	train	train	train
train	validation	train	train	train
train	train	validation	train	train
train	train	train	validation	train
train	train	train	train	validation

Five-fold cross-validation

"……generally when we say 'a model' we refer to a particular method for describing how some input data relates to what we are trying to predict. We don't generally refer to particular instances of that method as different models. "

E.g.
Model 1: linear regression;
Model 2: regression by second order polynomial;
……

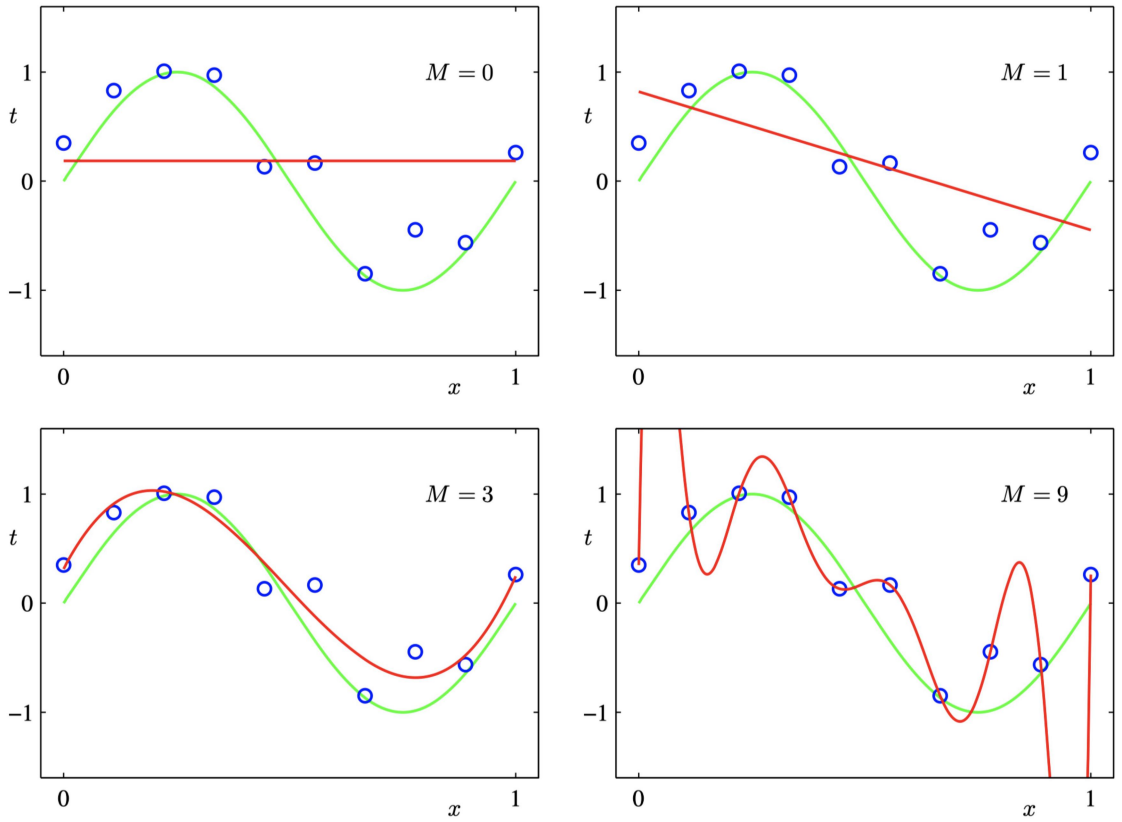
https://stats.stackexchange.com/questions/52274/how-to-choose-a-predictive-model-after-k-fold-cross-validation

Cross-validation: Choose the model with the smallest prediction error on the validation sets averaged over the five folds

Step 2: Picking a hypothesis space

- How to alleviate overfitting?

regularization Stem the coefficients from exploding



	$M=0$	$M=1$	$M=6$	$M=9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

Step 2: Picking a hypothesis space

- Linear regression
  - How to alleviate overfitting?

regularization — Stem the coefficients from exploding