

Machine Learning

國立中央大學

林省文

Data Mining

- 讓數據簡化而比較容易分析
 - 只留下最具有資訊的東西
- 在數據之中找出你想要的東西
- 找到數據中特別的Pattern
-

Data Mining 的方法

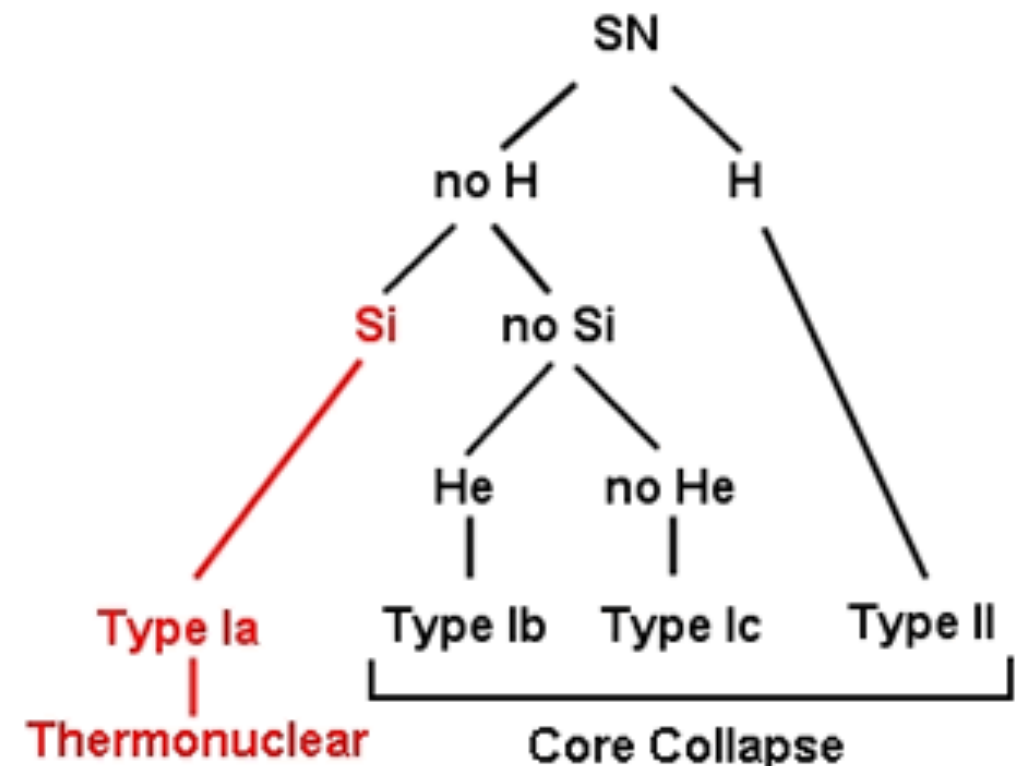
- 用眼睛看...(人力)
- 使用預先想好的模型選出想要的東西
- Machine Learning

用眼睛看.....

- Good: 眼睛的辨識能力很強, 不容易犯錯
- Bad: 很慢
- example: ZOONIVERSE <https://www.zooniverse.org>
 - Galaxy Zoo <http://www.galaxyzoo.org>
 - Planet Hunters <http://www.planethunters.org>
 -

使用模型

- Good: 很快很有效
- Bad: 沒有很完整知識, 很難建立好的模型
- example: 超新星的分類



Machine Learning

- 用電腦取代人力
- 訓練電腦，使電腦產生堪用的模型

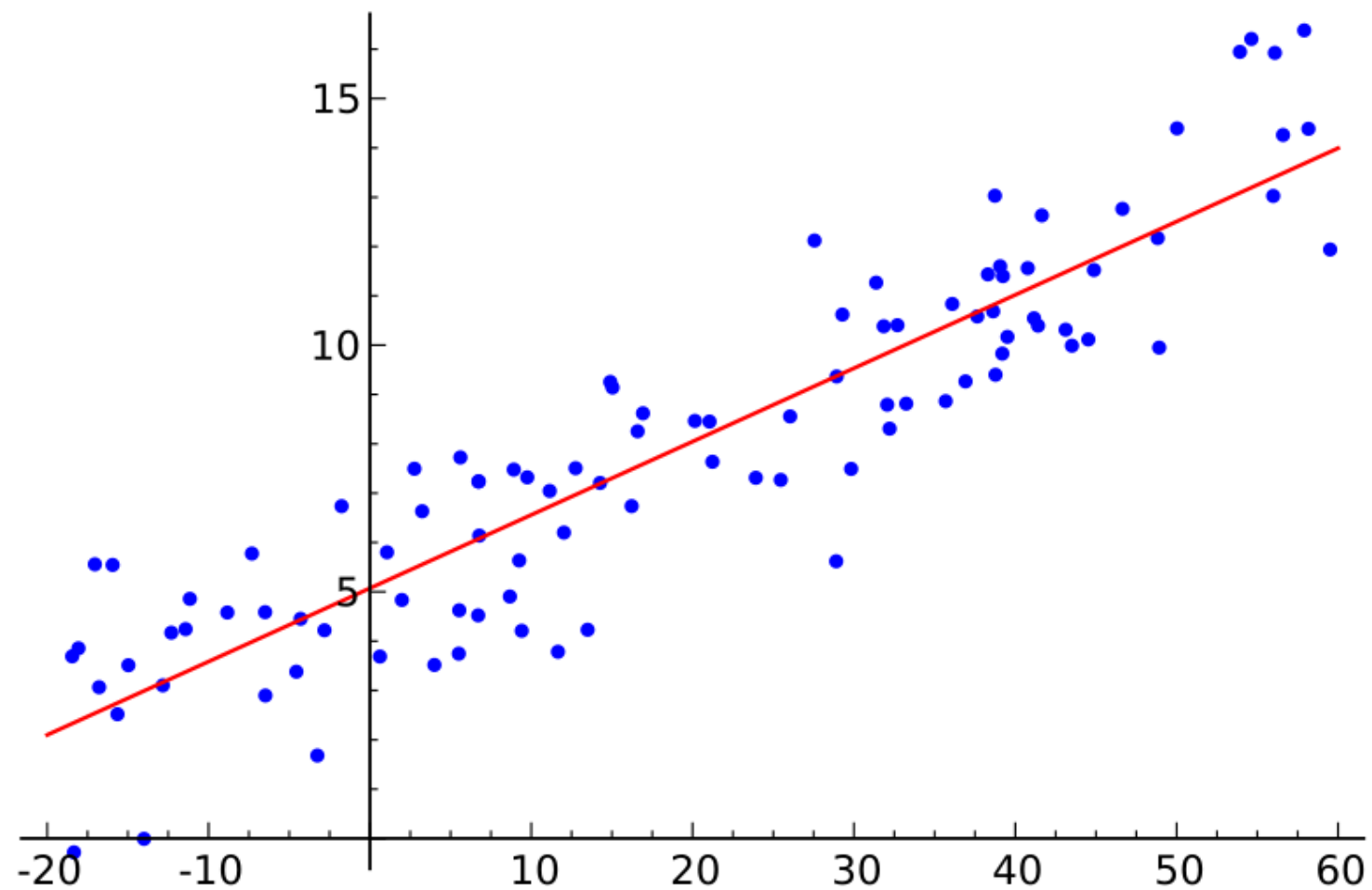
“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .”

–Tom M. Mitchell

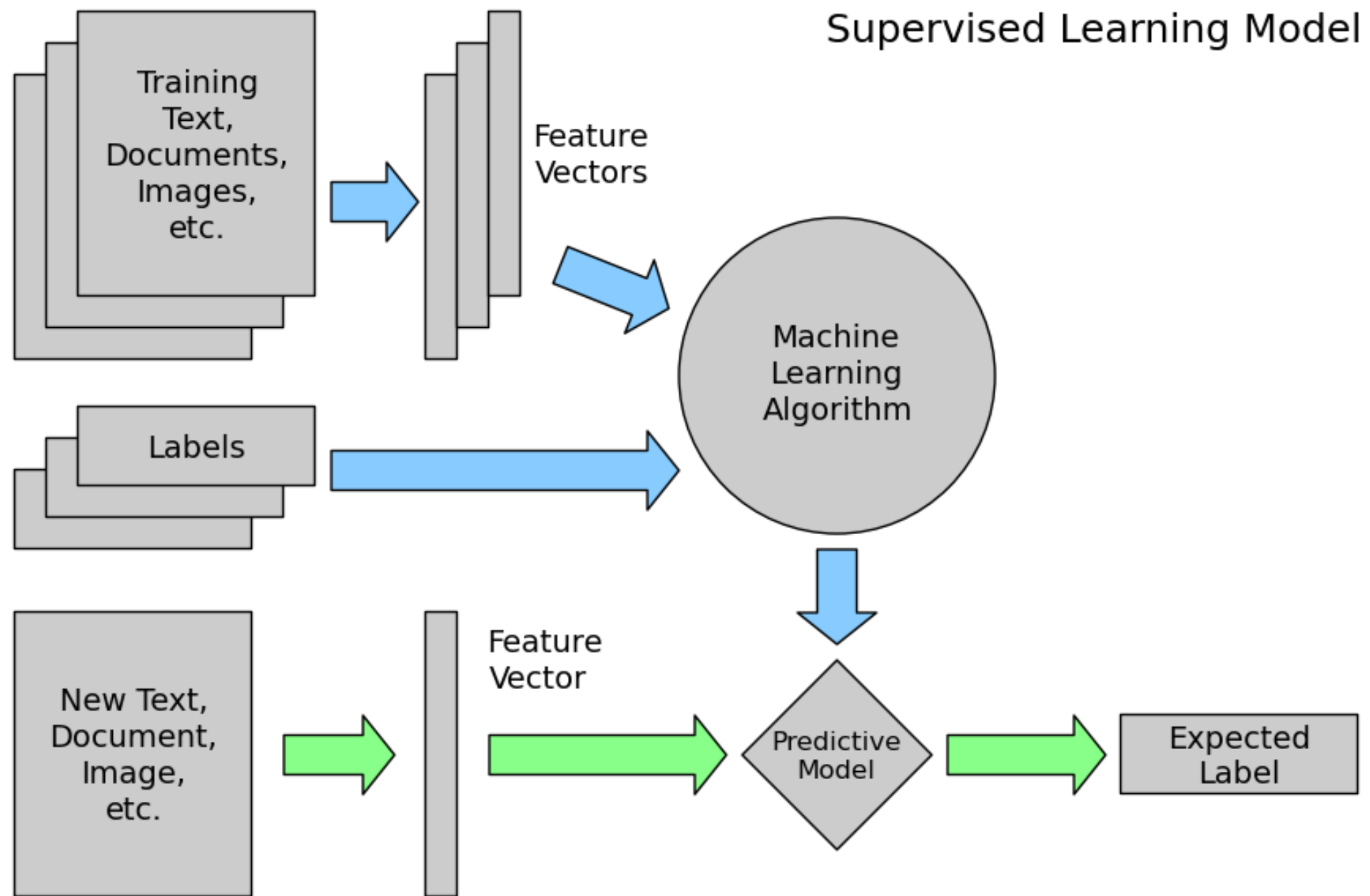
Fitting!!

Linear Regression Problem

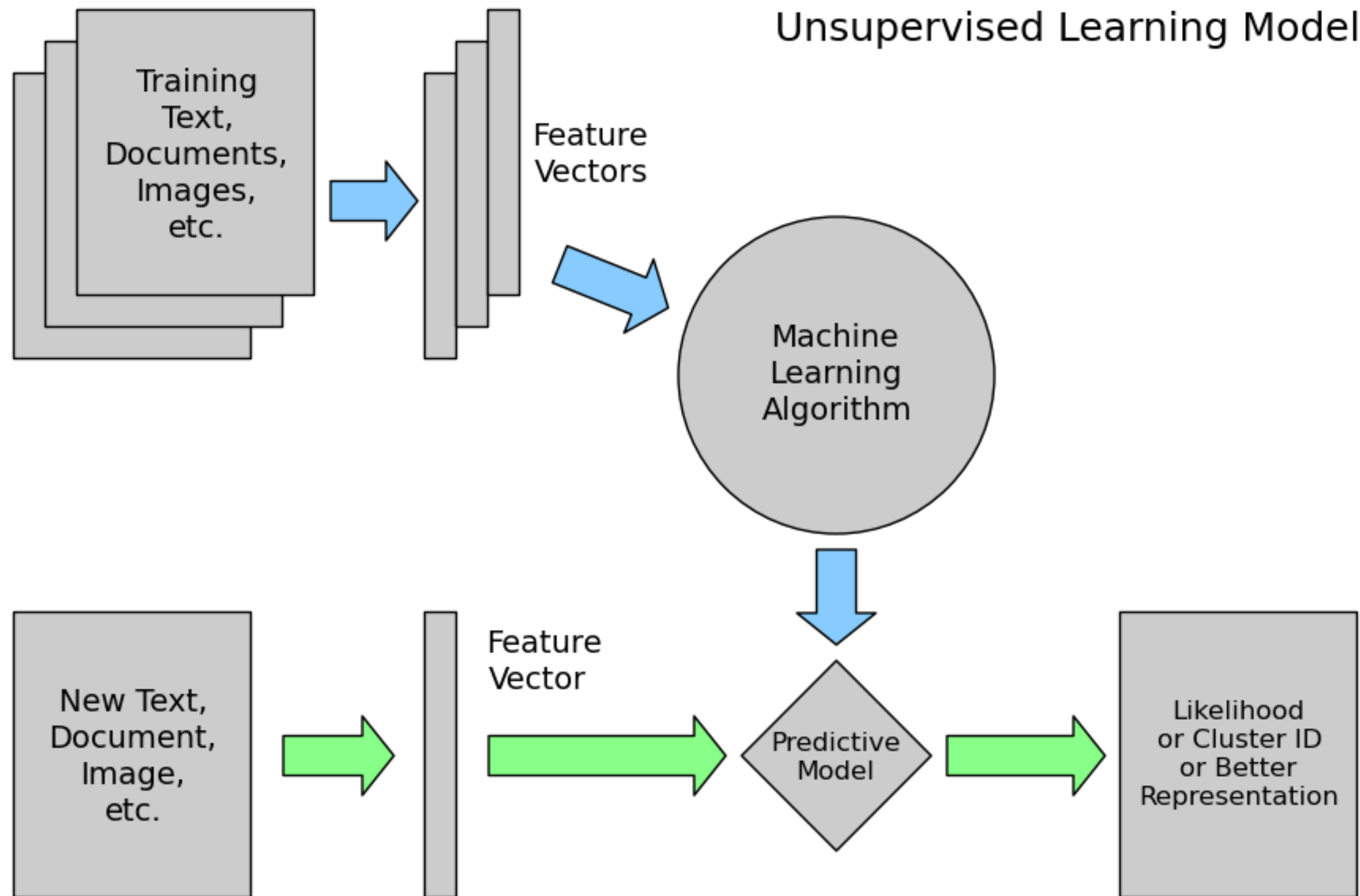
- $y = ax + b$



Supervised Learning



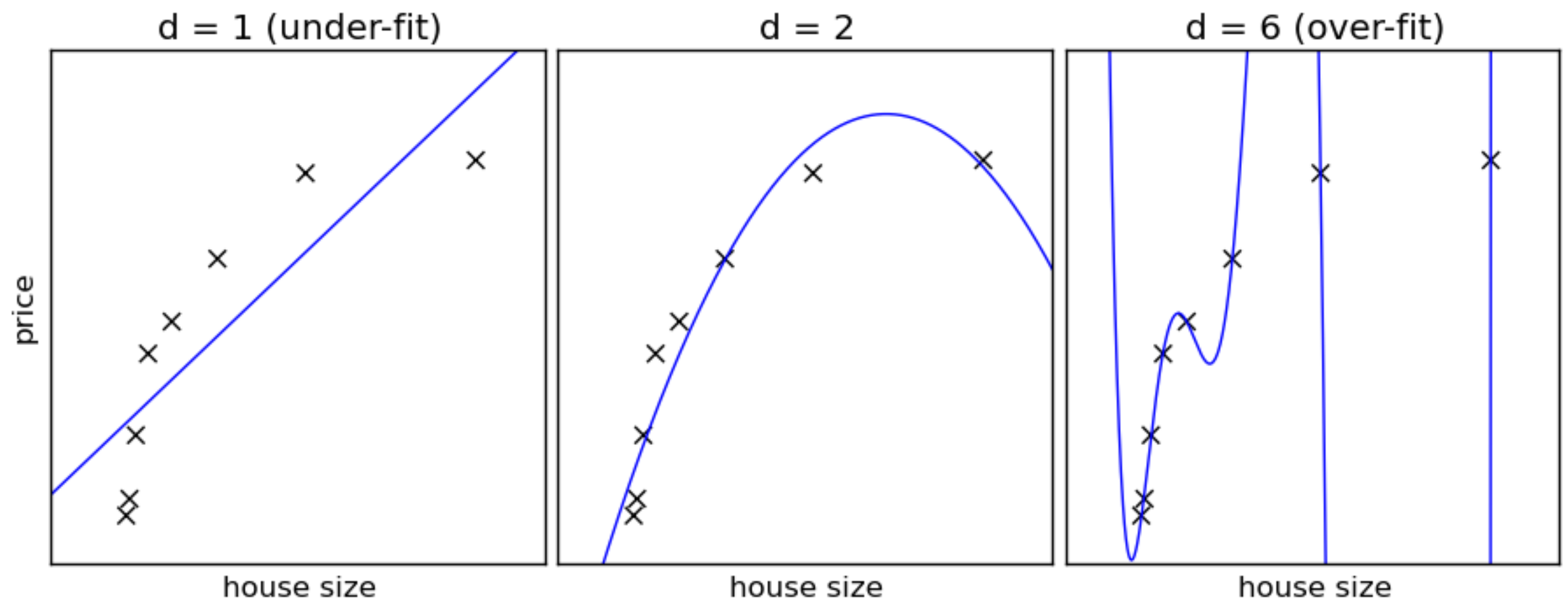
Unsupervised Learning



用途

- Supervised Learning: 分類, 回歸
- Unsupervised Learning: 類聚

Over Fitting



天文研究上的應用

- 過濾雜訊
- 分類天體 (利用顏色, 形狀等..)
- 分類變星 (利用光變曲線)
- 計算紅位移 (photo-z)
-

Reference

- http://www.astroml.org/sklearn_tutorial
- https://en.wikipedia.org/wiki/Machine_learning
- “DATA MINING AND MACHINE LEARNING IN ASTRONOMY”,
N. Ball and R. Brunner, International Journal of Modern
Physics D, 2010
- “Statistics, Data Mining, and Machine Learning in
Astronomy”, Z. Ivezić et al.

Decision Tree

