(Discussion to Tut 0 solns: Monday 1:30 Pm at F.C. Kohli)

Introduction to Machine Learning

Instructor: Prof. Ganesh Ramakrishnan

Lecture 2 - Supervised vs. Unsupervised Learning and Method of Least Squares

Supervised vs Unsupervised

Task: Given a basket of fresh fruits, you are asked to identify the *type* of each fruit in the basket
Eg: apple, banana, cherry, grape

Case: 1

- **Observations**: Size (parametrised using length, breadth, *etc.*), Shape, Color, weight, density, volume, second order features

redundant

Supervised vs Unsupervised

Task: Given a basket of fresh fruits, you are asked to identify the *type* of each fruit in the basket
Eg: apple, banana, cherry, grape

Case: 1

- **Observations**: Size (parametrised using length, breadth, *etc.*), Shape, Color . . . .
- **Train data**: Fruits in the basket **along with their labels**
- **Goal**: Develop ability to assign labels to new fruits based on **observations** made on them
- **Supervised Learning**: Achieve the **Goal** by learning from **Train data**

without labels

Case 2: no training data

- Given no **label** on each fruit, could you organize the basket by clubbing together fruits of the same type?
- *E.g.*:: Group together fruits that exhibit similar shape or color
- Groupings on the basis of *color*:

cherry, apple --

Case 2:

- Given no **label** on each fruit, could you organize the basket by clubbing together fruits of the same type?
- *E.g.*:: Group together fruits that exhibit similar shape or color
- Groupings on the basis of *color*:
  - **Red Color Group**: Apples and cheery
  - **Green Color Group**: Bananas and grapes
- Groupings on the basis of *size*:

Case 2:

- Given no **label** on each fruit, could you organize the basket by clubbing together fruits of the same type?
- *E.g.*:: Group together fruits that exhibit similar shape or color
- Groupings on the basis of *color*:
  - **Red Color Group**: Apples and cheery
  - **Green Color Group**: Bananas and grapes
- Groupings on the basis of *size*:
  - **Red color and big size**: Apple
  - **Red color and small size**: Cheery
  - **Green color and big Size**: Banana
  - **Green color and small Size**: Grapes
- This is **unsupervised learning**

Cross product of 2 observations

New observations could include taste

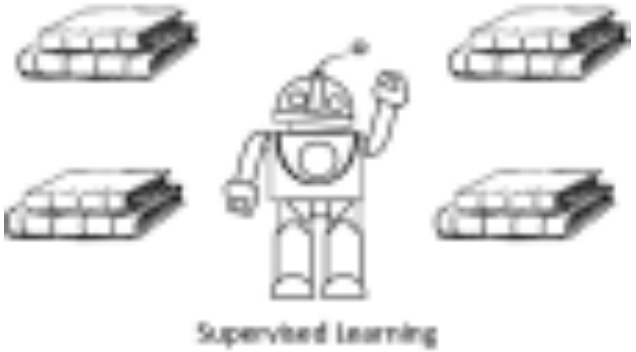
In case there is ambiguity (eg: big lemon & small orange)
↳ You can factor in more observations
↳ You can develop "models of probabilistic membership"

In fact, even Supervised learning could benefit from

- More observations
  - More attributes such as taste
  - More labeled examples
- Probabilistic models of uncertainty / certainty

1 - Pr(orange) Pr(orange)

Key Difference between Supervised and Unsupervised Learning



Another unsupervised setting:

Identifying a good subset of attributes from a somewhat redundant set

Redundant set

- {volume, wt, density, length, breadth ...}

- Supervised learning ⇒ Observed output is specified in the sample
- Unsupervised learning ⇒ Desired output is unobserved

length, volume, wt

I might find them less correlated for coconut

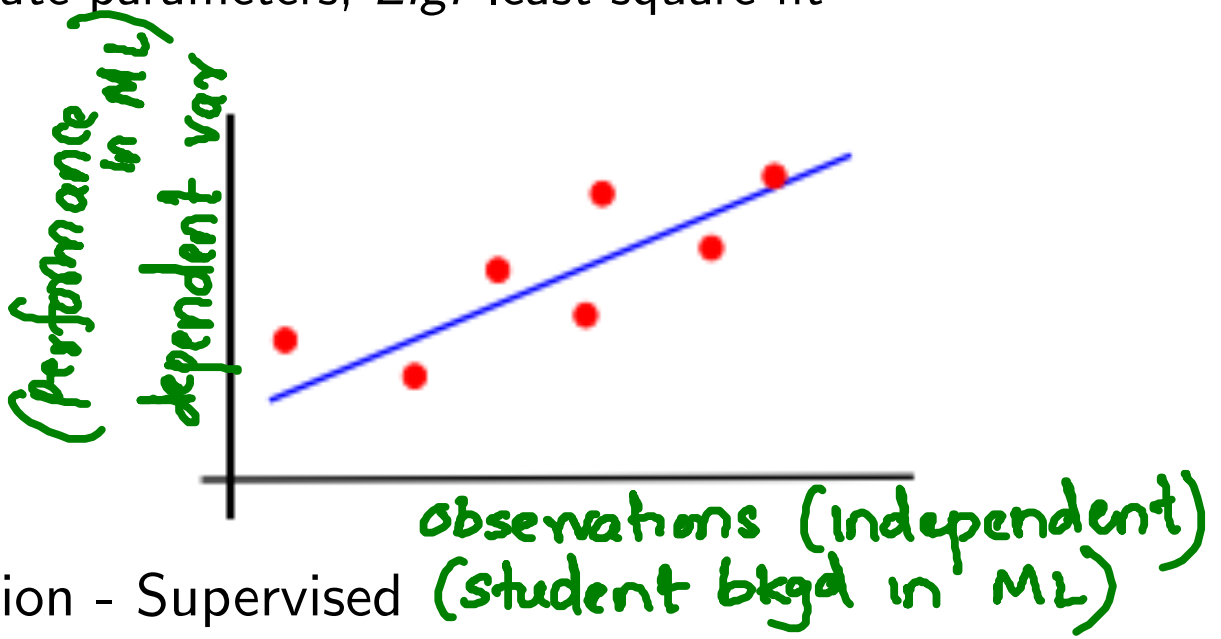
length, volume, wt

I might find them to be highly correlated for apple

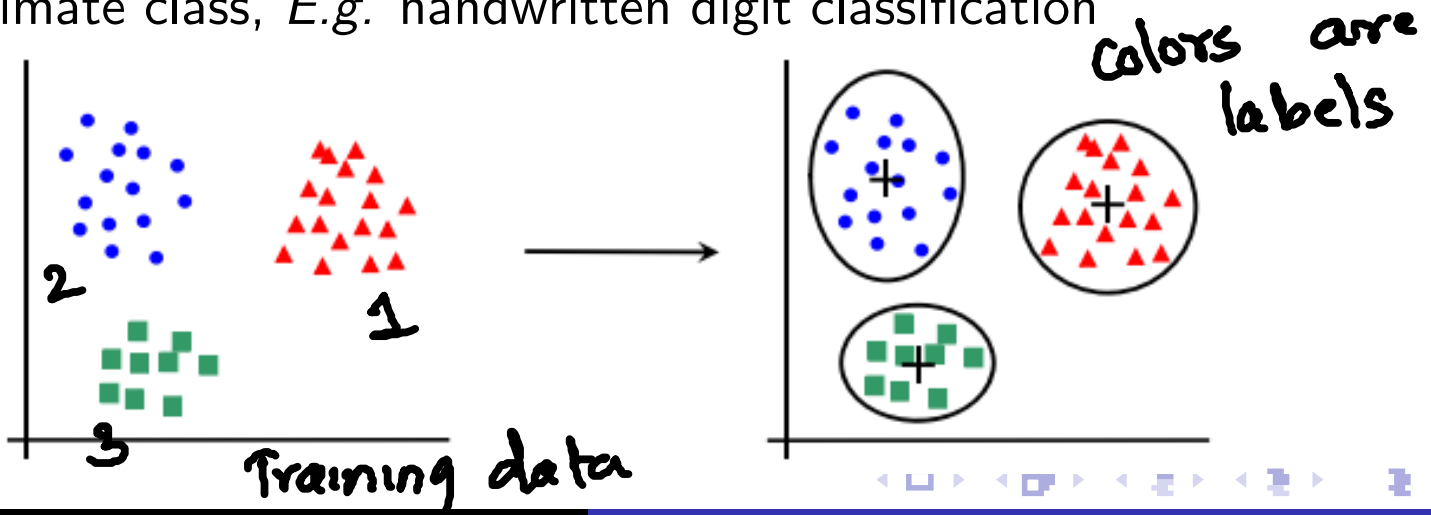
∴ Attribute subset selection could be benefitted through supervision!

Three Canonical Learning Settings

1. Regression - Supervised
   - Estimate parameters, *E.g.* least square fit



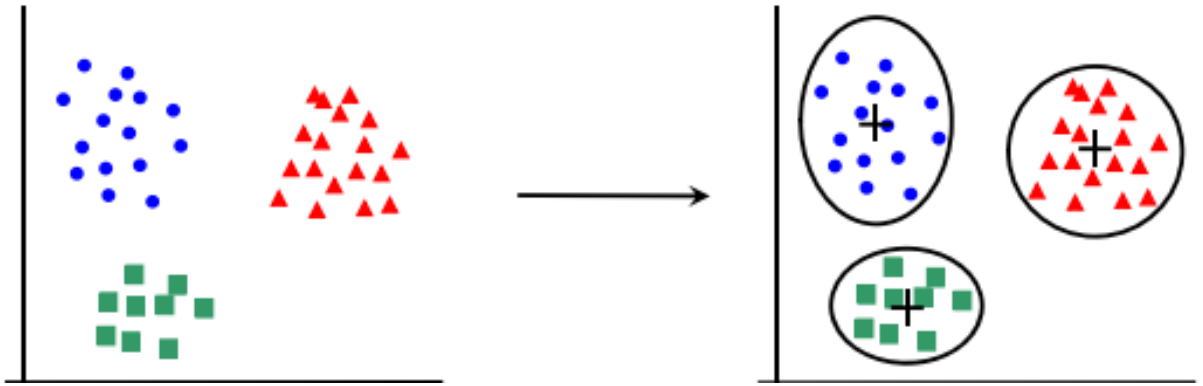
2. Classification - Supervised
   - Estimate class, *E.g.* handwritten digit classification



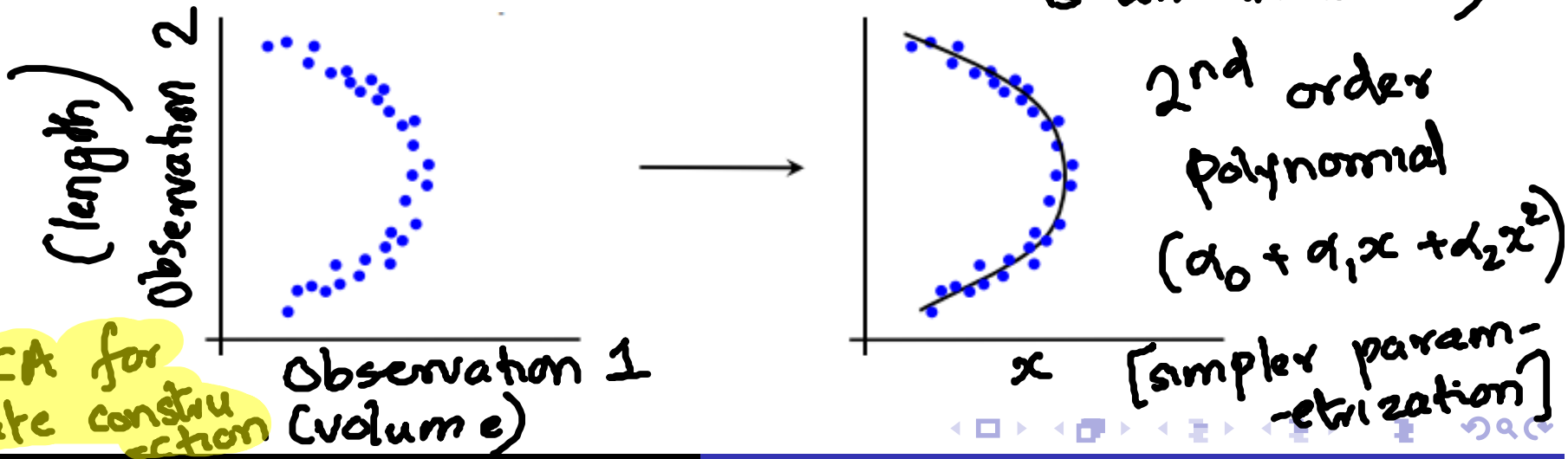
Three Canonical Learning Settings (contd.)

3 Unsupervised Learning - Model the data

- clustering (imagine colors dont exist)



- dimentionality reduction (attribute subset selection is an instance)



Eg: PCA for attribute construction

Supervised Learning: More formally

Functions F Training Data

$f : X \to Y$ $\{ (x^i, y^i) \in X \times Y \}$

i is example index

Supervised Learning

Functions F Training Data

$f : X \to Y$ $\{ (x^i, y^i) \in X * Y \}$

Eg: F = linear models Uncountably infinite

space of functions (F) in which to search for $\hat{f}$

Learning machine

LEARNING

find $\hat{f} \in \mathcal{F}$
s.t. $y_i \approx \hat{f}(x_i)$

PREDICTION

$y = \hat{f}(x)$ ⟸ New data x

Requirement for generalization (Prob theory, optimization)

Eg F = decision trees finite but too large

Training

$\{(x_i, y_i)\} \xrightarrow{\text{learn}} f \in \mathcal{F}$

Generalization

Desirable is

$f(x_{new}) \approx y_{new}$ → a new apple from Thailand

Probability

① For any x_{new} which comes close to distribution over x_i, I want $Pr(y_{new} | x_{new})$

$x_{new} \in P(x_i)$ — apples from Kashmir $\sim Pr(y_i | x_i)$

Optimization (one aspect)

② For any x_{new} which is "close" in some metric/distance to x_i's $(\leq d(x_{new}, x_i) < \theta)$

I would like $y_{new} = \underset{d_x(y_i, y_{new})}{\text{closest}} y_i$ as per

Optimisation in clustering (unsupervised)

Find groups such that

(intra group) ↳ Similarity within each group is MAXIMIZED

(inter group) ↳ Similarity across groups is MINIMIZED

So far....

- **Machine Learning in General**
  - Supervised Learning
  - Unsupervised Learning
  - Applications and examples
- **Canonical Learning Problems**
  - Regression Supervised
  - Classification Supervised
  - Unsupervised modeling of data

We will develop many tools using this canonical example

① Optimization

② Generalization using optimization such as regularization, Bayesian inference

③ Parametric vs non-param learning

④ Non-linear models via the kernel trick & duality

⑤ Dimensionality reduction