The Machine Learning Process

Machine learning problems involve multiple steps

1) Data collection

- camera
- database

2) Preprocesing

- common size, center
- reject poor quality images

★ Time consuming!

Important!

3) Feature extraction

☆What attributes are relevant?

→ Only select some of the features

- Collect measurements in a vector

4) Collect training data

☆(The raw data after being processed)

☆(Divide the training and test set)

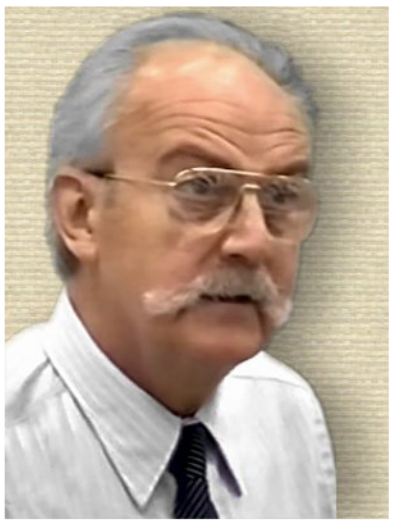
Barry Gauss Dave

Labels

→ known examples

5) Apply a model - captures salient aspects of the data

- face identity represented by differences in landmarks relative to the mean
- movie ratings are a weighted combination of nominal taste profiles



"All models are wrong, but some are useful" George E. P. Box (1919 - 2013)

Two general categories of problems:

1) Supervised learning ☆ (Have labeled source data)
   - examples with correct outcomes
   - train model with examples
   - apply trained model to new data
   Ex: classification  vs 

2) Unsupervised learning
   - no training data/examples

   Ex: clustering ☆ Characterize the patterns of data