Special Module on Intelligent Information Processing

Dayalbagh Educational Institute
(DEI)
Dayalbagh Agra

CSM 802

Indian Institute of Technology Delhi
(IITD)
New Delhi

SIV 895

What is Machine Learning

"Learning is any process by which a system improves performance from experience." Herbert Simon

Definition by Tom Mitchell (1998):

Machine Learning is the study of algorithms that

- improve their performance P
- at some task T
- with experience E.

A well-defined learning task is given by <P, T, E>.

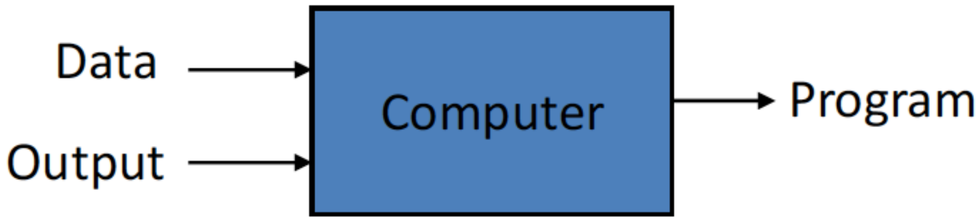
What is Machine Learning

"Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed." - Arthur Samuel (1959)

What is Machine Learning

Traditional Programming

Machine Learning

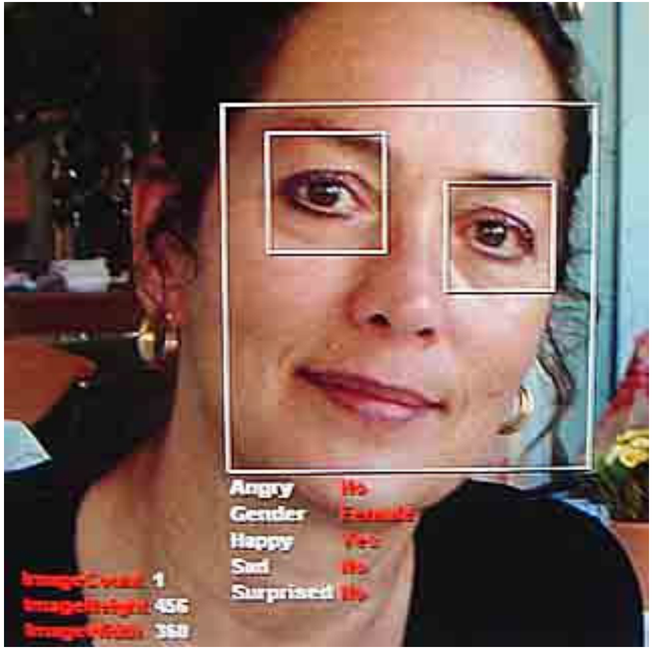


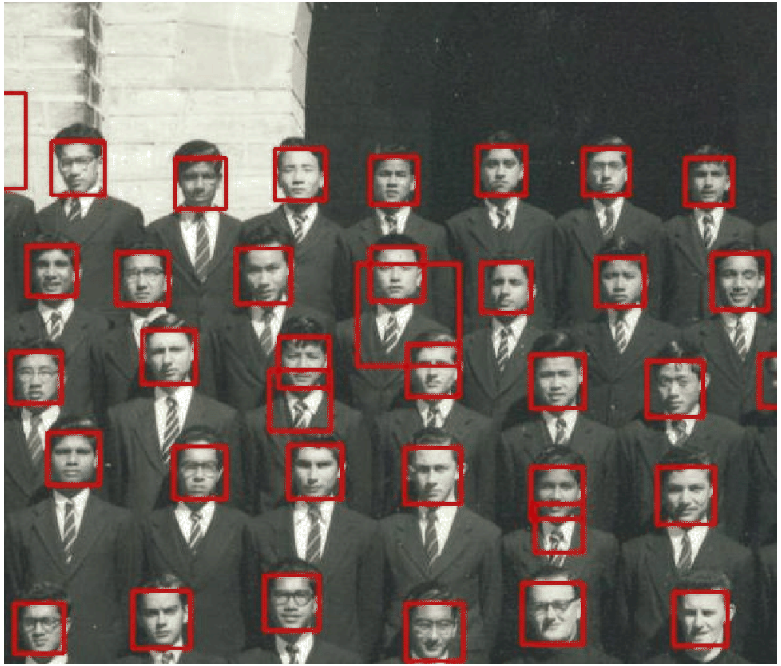
Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Where is it used

- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Generating patterns:
  - Generating images or motion sequences
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates

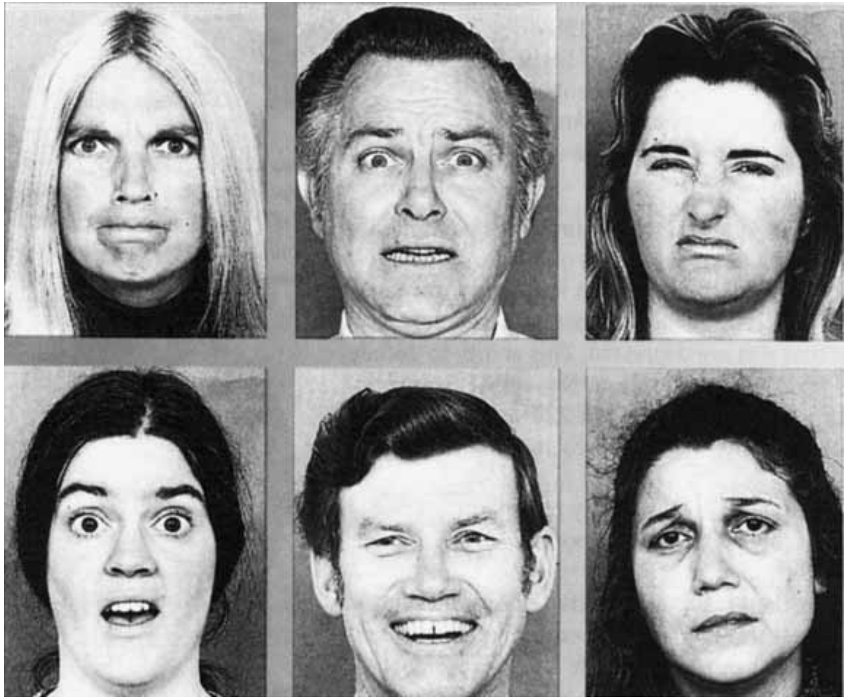
Face Detection/Recognition





Facial Expression Recognition

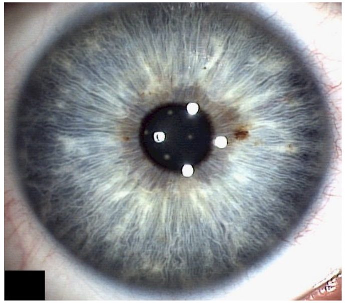
Anger Fear Disgust



Surprise Happiness Sadness

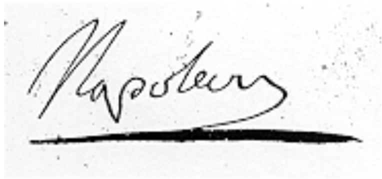
Object Identification/Recognition

Fingerprint

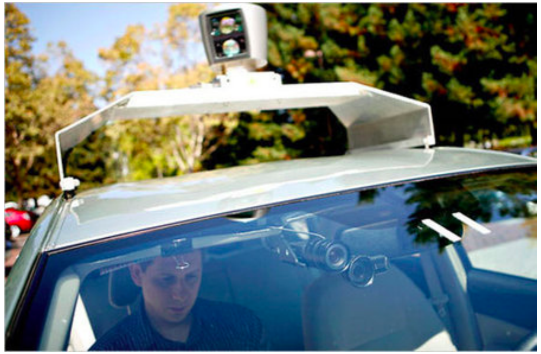


Iris

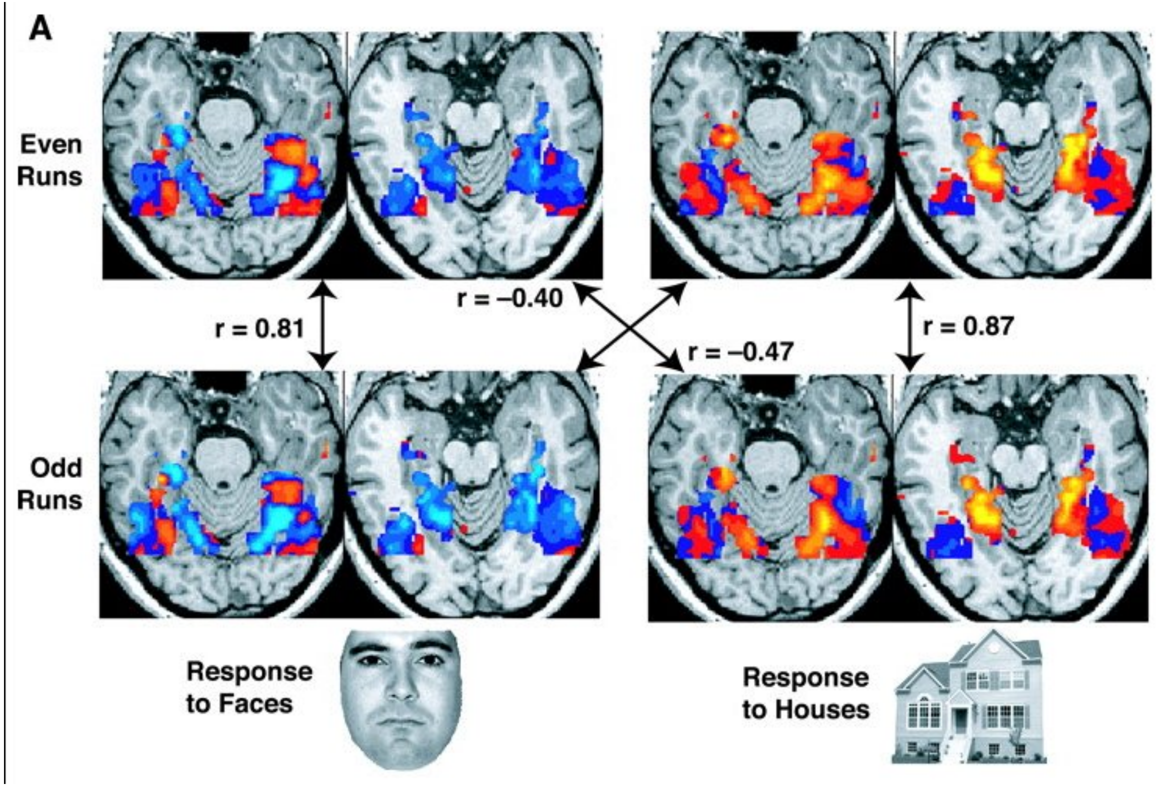
Face



Autonomous/Assisted Cars



Medical Images



How are things learnt

- Memorization
  - Accumulation of individual facts
  - Limited by
    - Time to observe facts
    - Memory to store facts

Declarative knowledge

- Generalization
  - Deduce new facts from old facts
  - Limited by accuracy of deduction process
    - Essentially a predictive activity
    - Assumes that the past predicts the future

Imperative knowledge

- Interested in extending to programs that can infer useful information from **implicit** patterns in data

Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Basic Paradigm

Observe set of examples: training set

Infer something about process that generated that data

Use inference to make predictions about previously unseen data: test set

Types of learning:

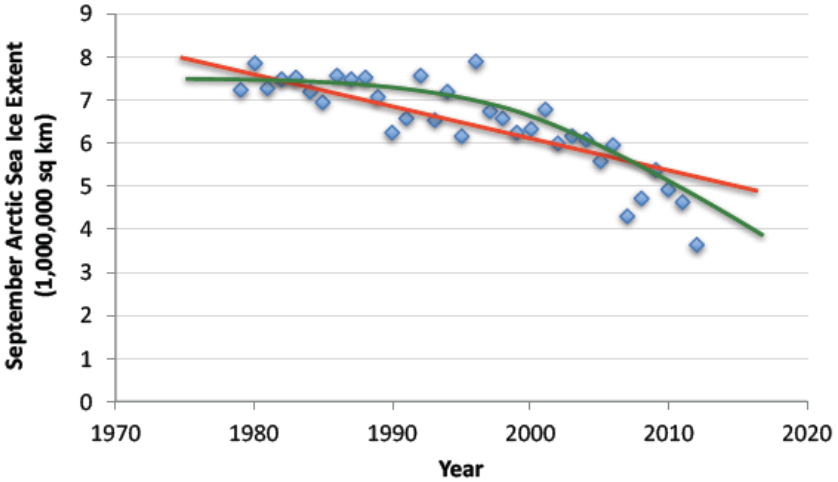
Supervised: given a set of features/label pairs, find a rule that predicts the label association with unseen data

Unsupervised: given a set of feature vectors (without labels), find natural groups or clusters (create labels for groups)

Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Supervised Learning: Regression

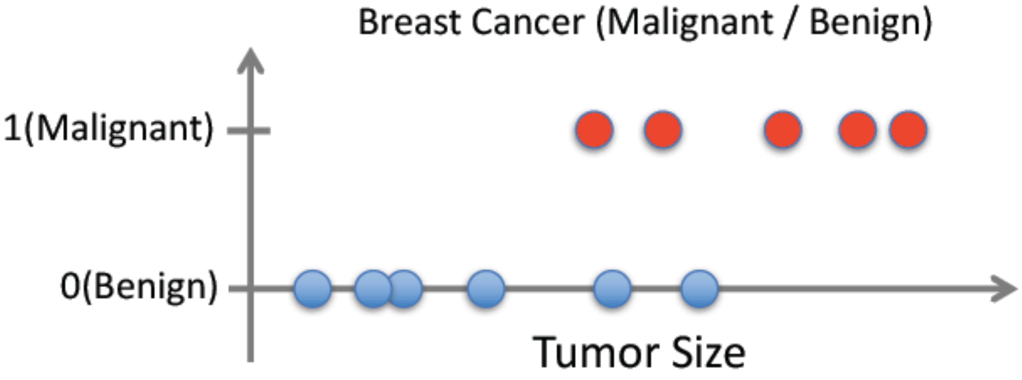
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



Adapted from source:Introduction to Machine Learning by Eric Eaton

Supervised Learning: Classification

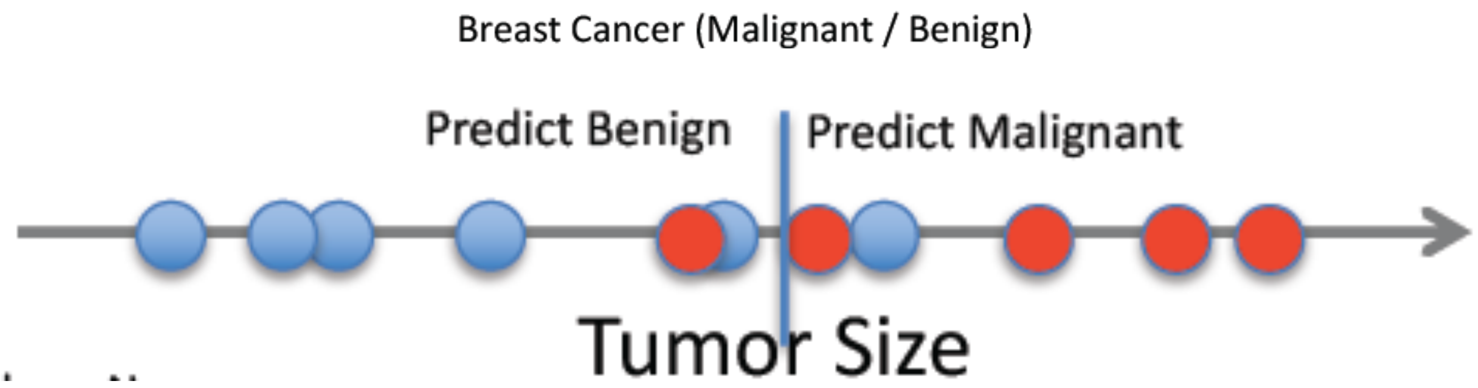
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification



Adapted from source:Introduction to Machine Learning by Eric Eaton

Supervised Learning: Classification

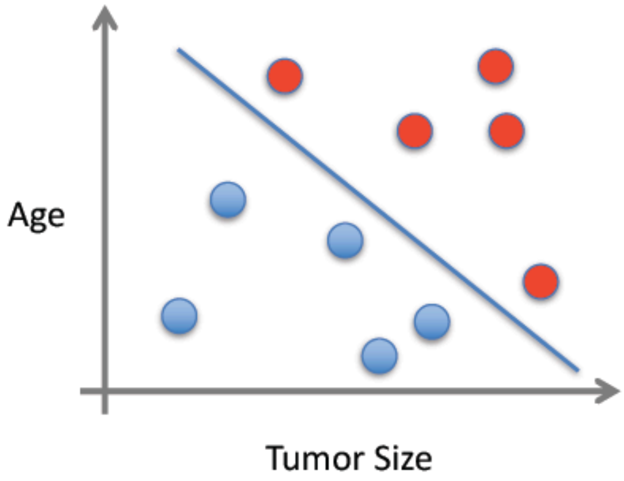
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification



Adapted from source:Introduction to Machine Learning by Eric Eaton

Supervised Learning: Classification

- x can be multi-dimensional
  – Each dimension corresponds to an attribute

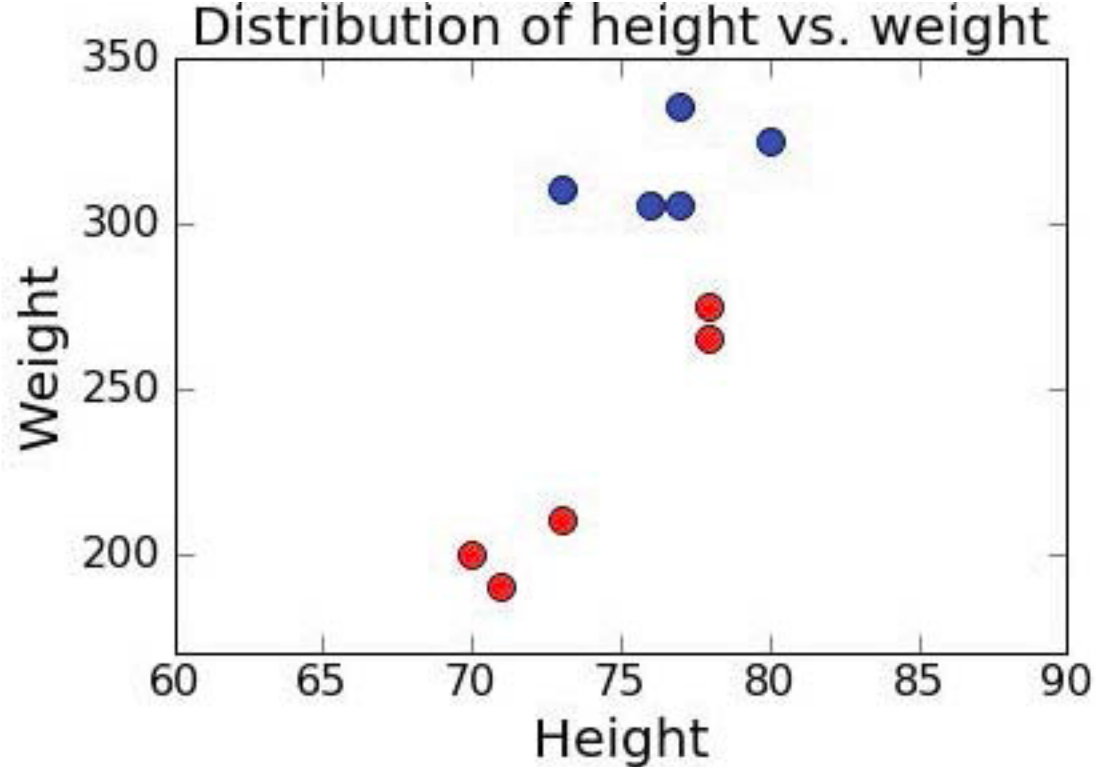


- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

…

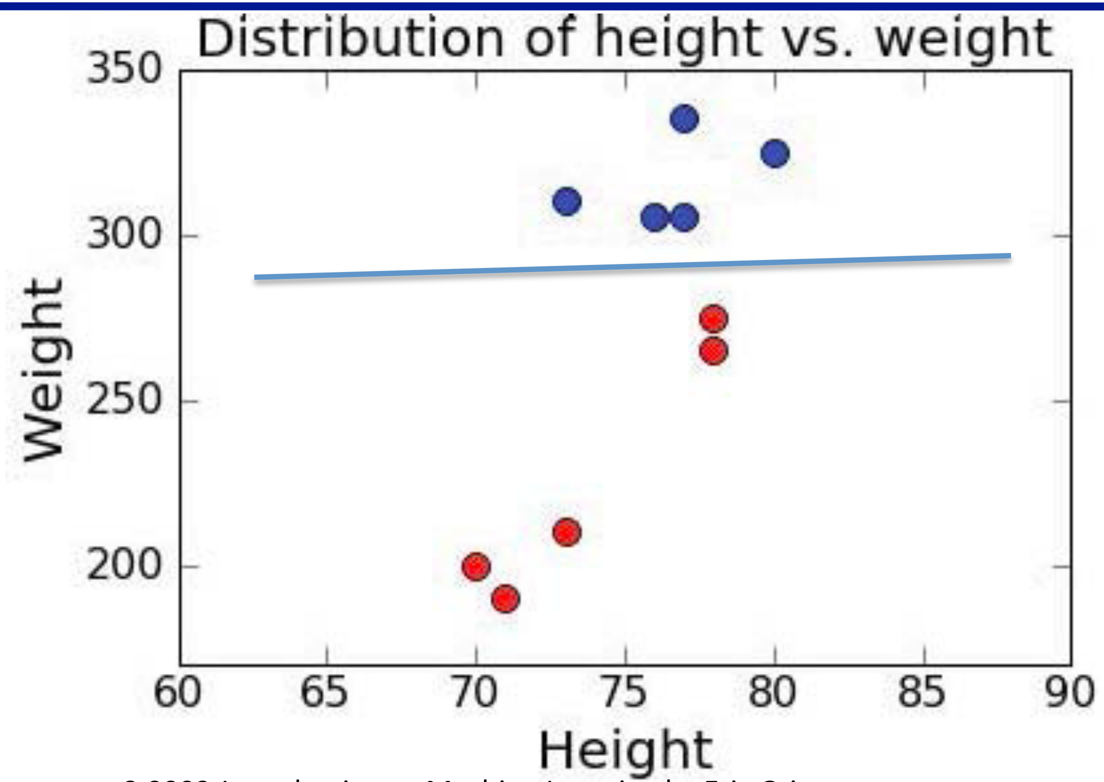
Adapted from source:Introduction to Machine Learning by Eric Eaton

Supervised Learning: Classification



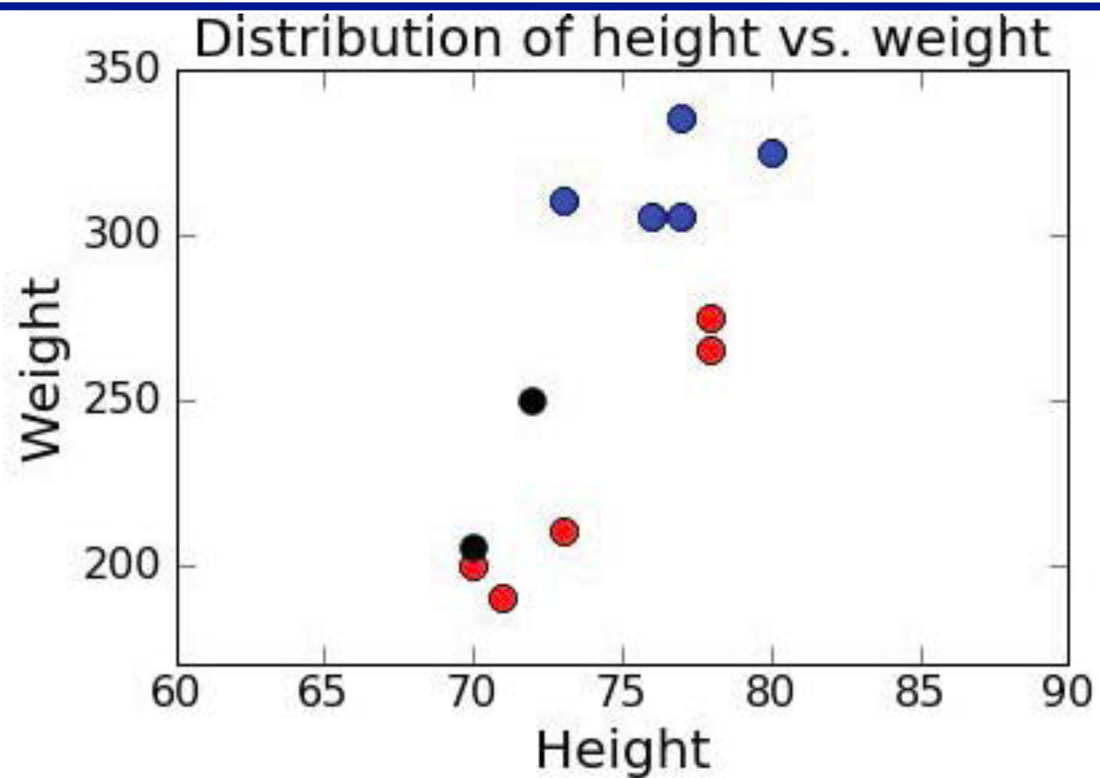
Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Supervised Learning: Classification



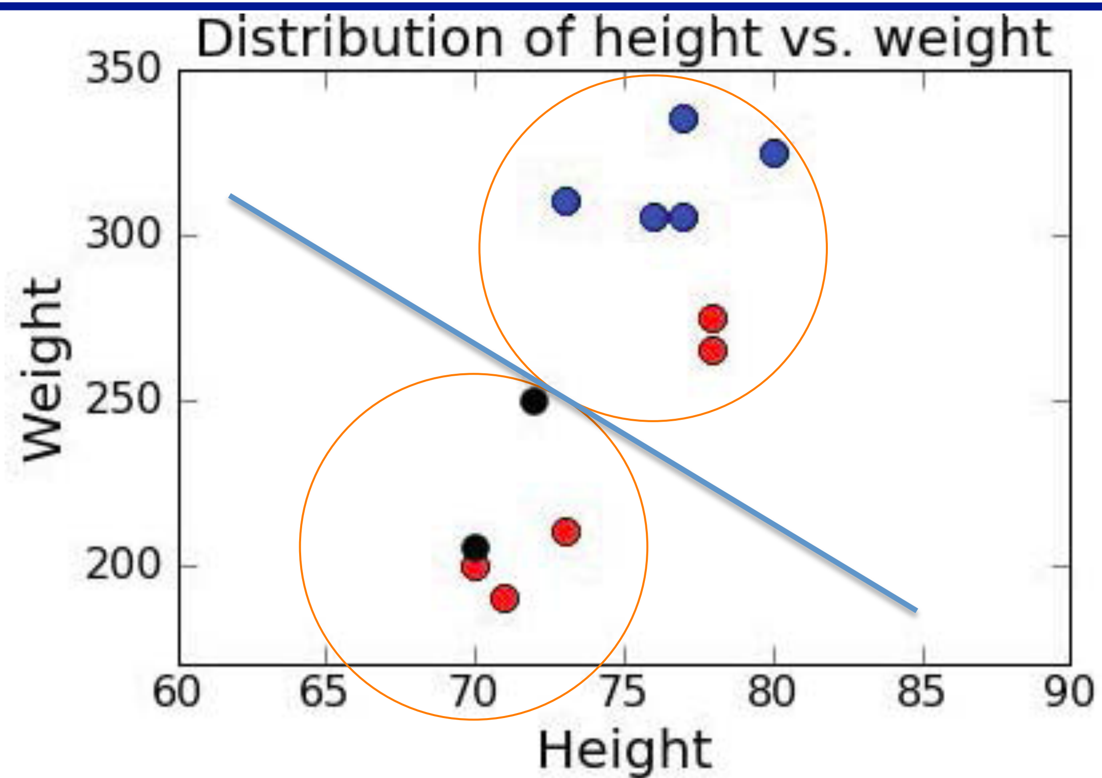
Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Supervised Learning: Classification



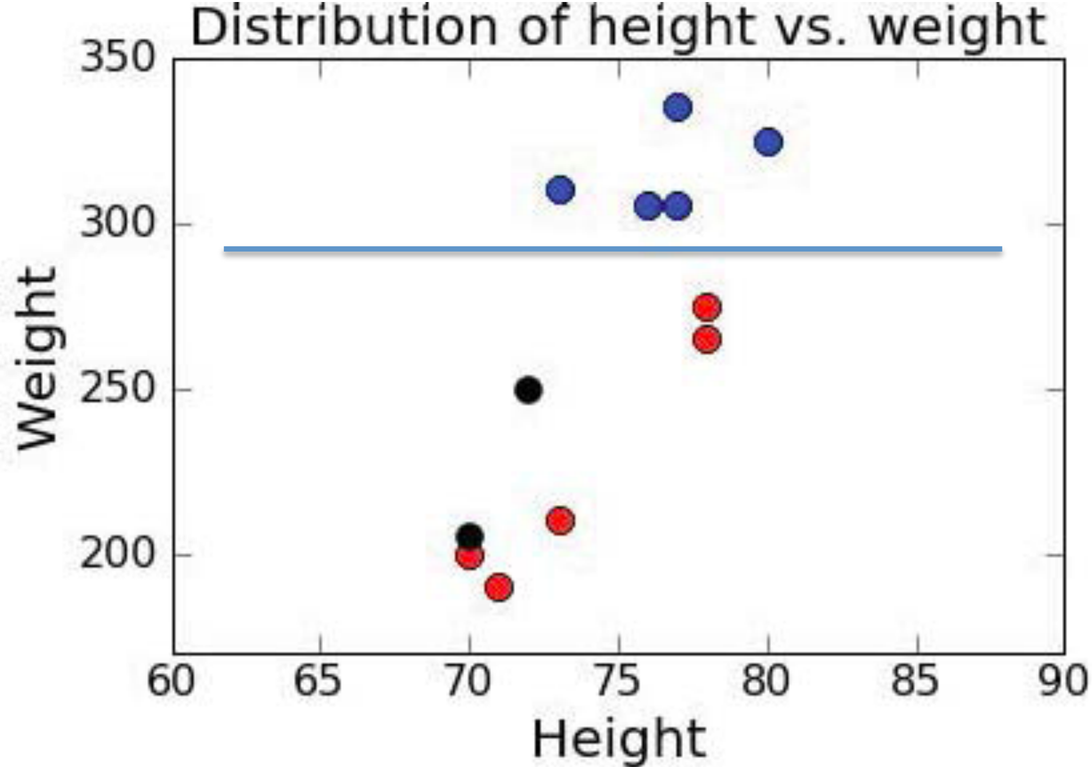
Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Supervised Learning: Classification



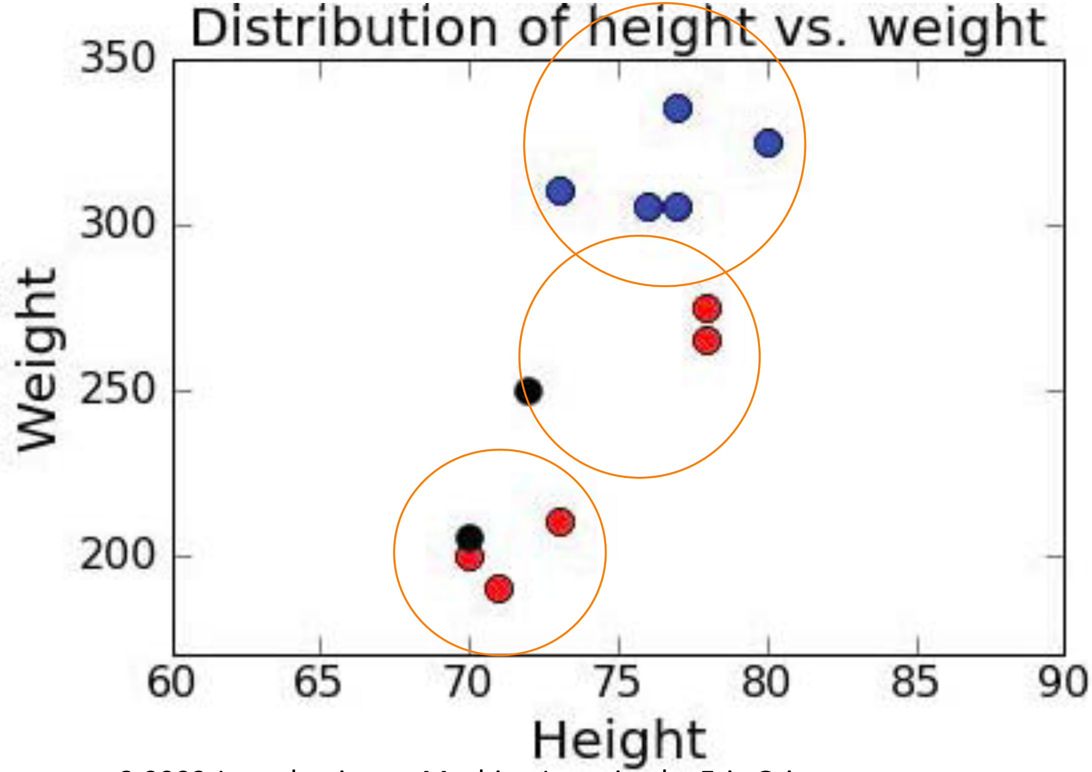
Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Supervised Learning: Classification



Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

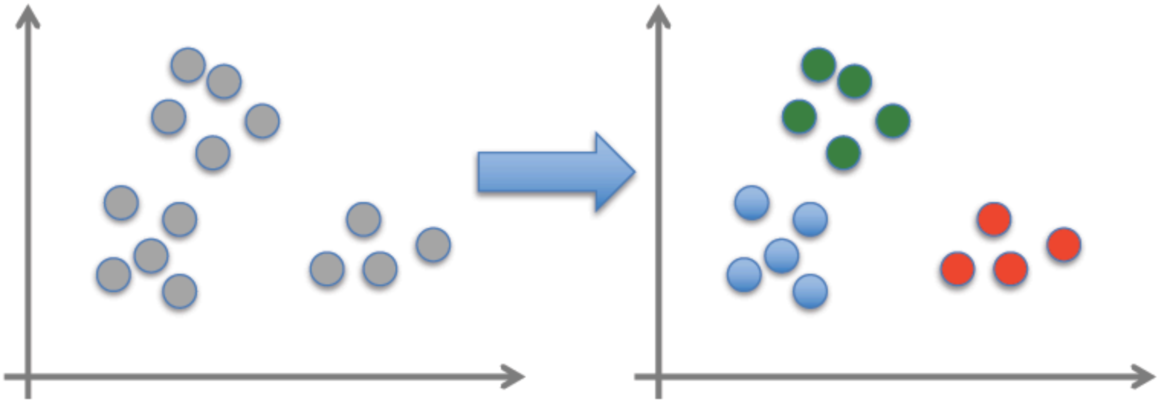
Supervised Learning: Classification



Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

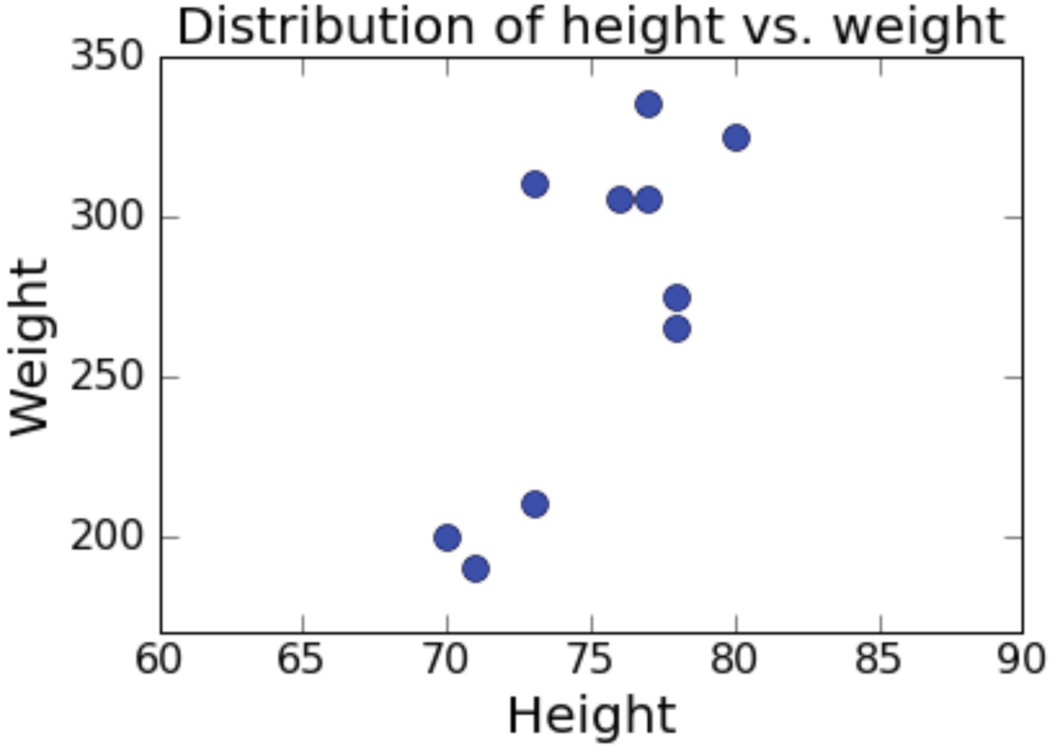
Unlabeled Data

- Given x_1, x_2, …, x_n (without labels)
- Output hidden structure behind the x's
  - E.g., clustering



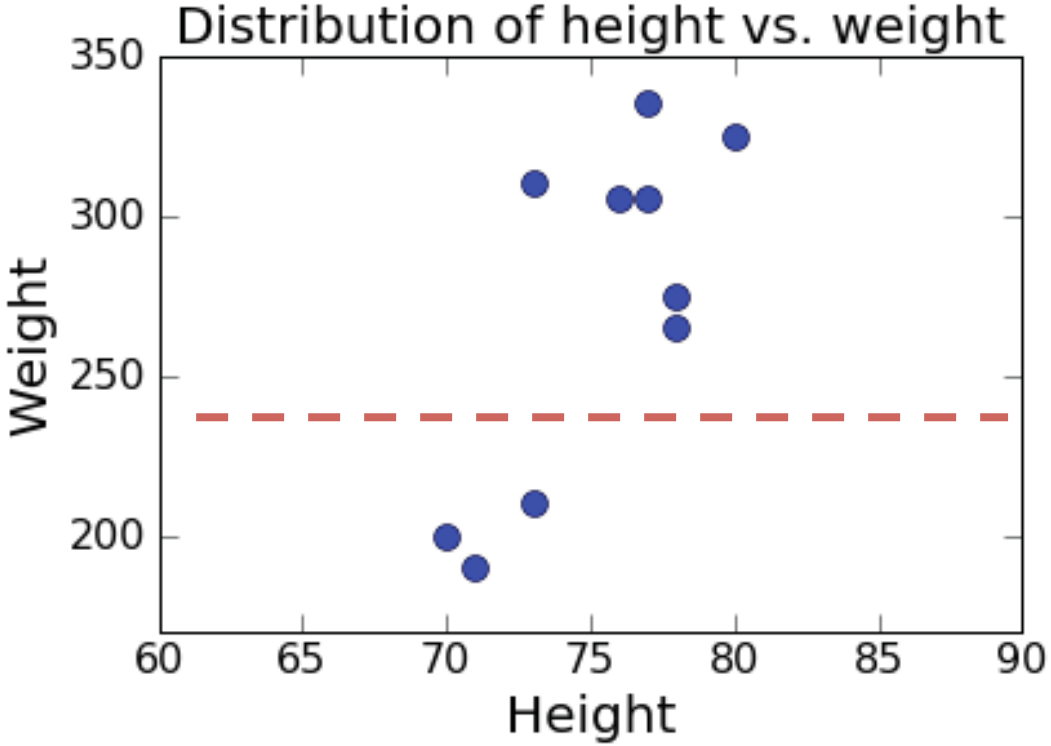
Adapted from source:Introduction to Machine Learning by Eric Eaton

Unlabeled Data



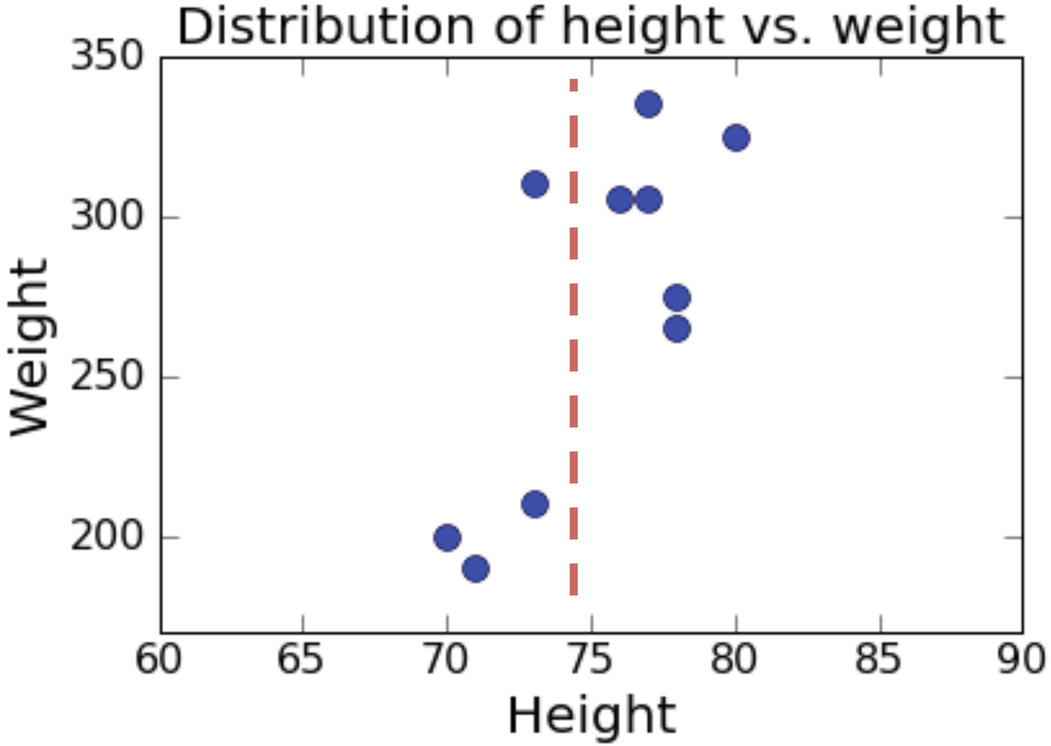
Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Unlabeled Data



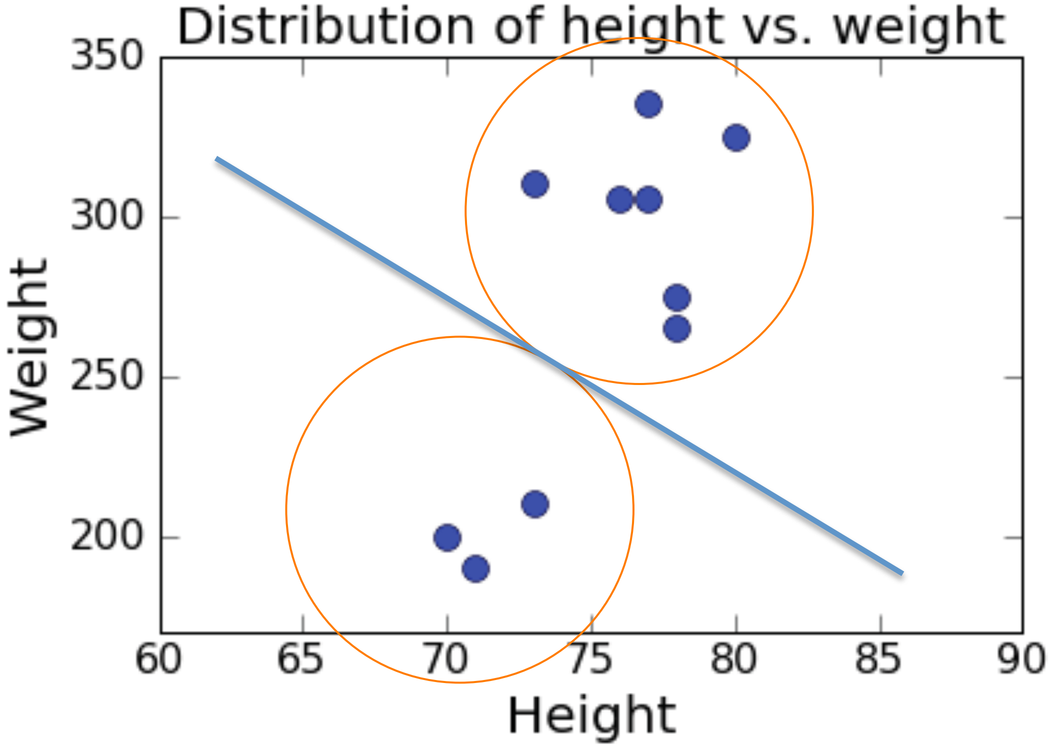
Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Unlabeled Data



Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Unlabeled Data



Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Features

- Feature engineering
  - Represent examples by feature vectors that will facilitate generalization
  - Suppose I want to use 100 examples from past to predict, at the start of the subject, which students will get an A
  - Some features surely helpful, e.g., GPA, prior programming experience (not a perfect predictor)
  - Others might cause me to overfit, e.g., birth month, eye color
- Want to maximize ratio of useful input to irrelevant input
  - Signal-to-Noise Ratio (SNR)

Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Features

- Feature engineering:
  - Deciding which features to include and which are merely adding noise to classifier
  - Defining how to measure distances between training examples (and ultimately between classifiers and new instances)
  - Deciding how to weight relative importance of different dimensions of feature vector, which impacts definitionof distance

Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

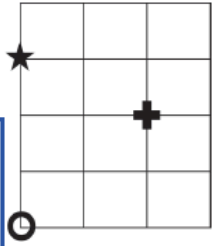
Distance

Minkowski Metric

$$dist(X1, X2, p) = \left(\sum_{k=1}^{len} abs(X1_k - X2_k)^p\right)^{1/p}$$

p = 1: Manhattan Distance
p = 2: Euclidean Distance

Need to measure distances between feature vectors



Typically use Euclidean metric; Manhattan may be appropriate if different dimensions are not comparable

Is circle closer to star or cross?

- Euclidean distance
  - Cross – 2.8
  - Star – 3
- Manhattan Distance
  - Cross – 4
  - Star - 3

Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Confusion Matrix

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

- There are two possible predicted classes: "yes" and "no". If we were predicting the presence of a disease, for example, "yes" would mean they have the disease, and "no" would mean they don't have the disease.
- The classifier made a total of 165 predictions (e.g., 165 patients were being tested for the presence of that disease).
- Out of those 165 cases, the classifier predicted "yes" 110 times, and "no" 55 times.
- In reality, 105 patients in the sample have the disease, and 60 patients do not.

Adapted from source: https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

Confusion Matrix

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

- **true positives (TP):** These are cases in which we predicted yes (they have the disease), and they do have the disease.
- **true negatives (TN):** We predicted no, and they don't have the disease.
- **false positives (FP):** We predicted yes, but they don't actually have the disease. (Also known as a "Type I error.")
- **false negatives (FN):** We predicted no, but they actually do have the disease. (Also known as a "Type II error.")

Adapted from source: https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

Accuracy

$$accuracy = \frac{true\ positive + true\ negative}{true\ positive + true\ negative + false\ positive + false\ negative}$$

Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Other Measures

$$sensitivity = \frac{true\ positive}{true\ positive + false\ negative}$$

$$specificity = \frac{true\ negative}{true\ negative + false\ positive}$$

Adapted from source:6-0002-Introduction to Machine Learning by Eric Grimson

Minimum Distance Classifier

A Decision Theoretic Approach

Let $x = (x_1, x_2, \ldots, x_n)^T$ for W pattern classes $\omega_1, \omega_2, \ldots, \omega_W$

$$d_i(x) > d_j(x) \quad j = 1,2,\ldots,W; j \neq i$$

- In other words, an unknown pattern **x** is said to belong to the *i*th pattern class if, upon substitution of **x** into all decision functions, $d_i(x)$ yields the largest numerical value.

Adapted from source: Digital Image Processing Gonzalez and Woods

Minimum Distance Classifier

- Suppose that we define the prototype of each pattern class to be the mean vector of the patterns of that class: $m_j = \frac{1}{N_j} \sum_{x \in \omega_j} x_j \quad j = 1,2,\ldots,W$
- We then assign **x** to class ω_i if $D_i(\mathbf{x})$ is the smallest distance. $D_j(x) = \|x - m_j\|$

Adapted from source: Digital Image Processing Gonzalez and Woods

Minimum Distance Classifier

$$d_j(x) = x^T m_j - \frac{1}{2} m_j^T m_j \quad j = 1, 2, \ldots, W$$

assign **x** to class ω_i if $d_i(\mathbf{x})$ is the largest numerical value.

Adapted from source: Digital Image Processing Gonzalez and Woods

Minimum Distance Classifier

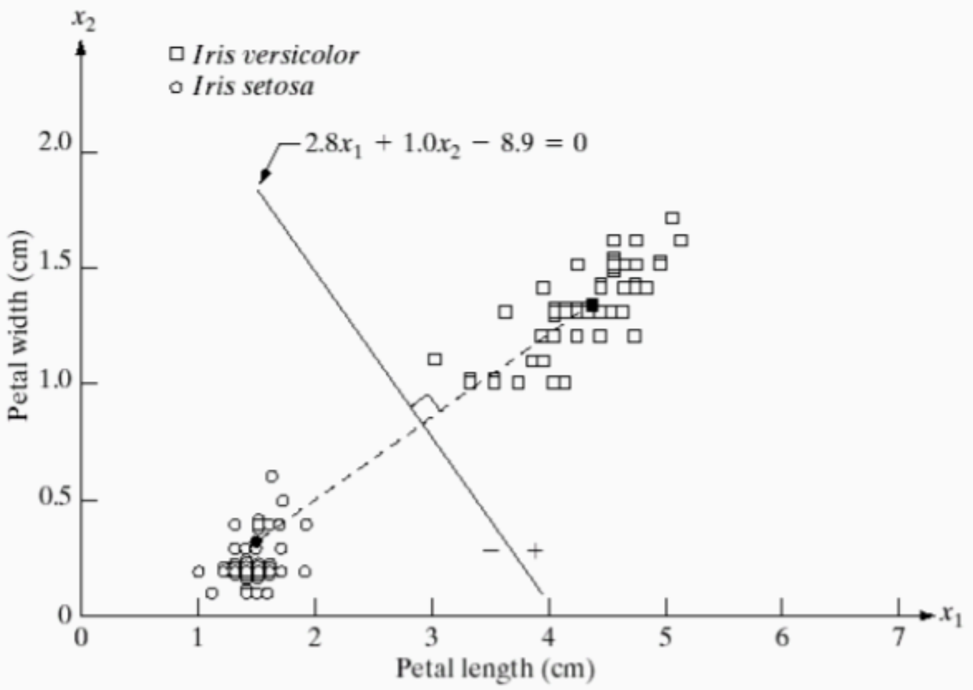


FIGURE 12.6 Decision boundary of minimum distance classifier for the classes of *Iris versicolor* and *Iris setosa*. The dark dot and square are the means.

Adapted from source: Digital Image Processing Gonzalez and Woods

Home Assignment

Consider two class scenario with 2D features (x1,x2). Is the minimum distance classifier boundary perpendicular to the line joining the two means (prototypes)? Justify your answer.