

Inductive learning hypothesis: any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over any other unobserved examples.

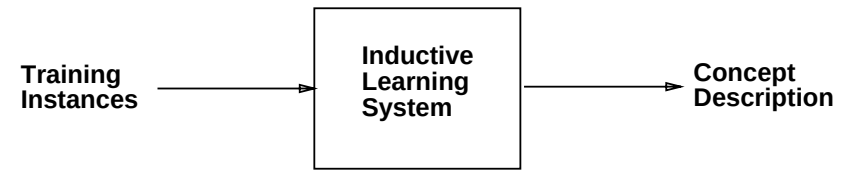
Assumptions for Inductive Learning Algorithms:

- The training sample represents the population
- The input features permit discrimination

Slide CS472 – Machine Learning 9

Inductive Learning

System tries to induce a general rule from a set of observed instances.



The *hypothesis* produced is sometimes called the *concept description* — essentially a program that can be used to classify subsequent instances.

Slide CS472 – Machine Learning 10

***k*-nearest neighbor**

Also called instance-based Learning; case-based learning.

A : set of features/attributes, $A_1, \dots, A_n$ that describe the problem

$x = x_{a_1}x_{a_2} \dots x_{a_n}$, where x_{a_i} is the value of feature A_i in example x

$$f(x) : x \rightarrow c \in C = \{c_1, \dots, c_m\}$$

The *case base* is the set of training examples
 $(x_1, f(x_1)), (x_2, f(x_2)), \dots$

Slide CS472 – Machine Learning 11

***k*-nearest neighbor algorithm for computing $f(x)$:**

1. Compare new example, x , to each case, y , in the case base and calculate for each pair:

$$\text{sim}(x, y) = \sum_{i=1}^n \text{match}(x_{a_i}, y_{a_i})$$

where $\text{match}(a, b)$ is a function that returns 1 if a and b are equal and 0 otherwise.

2. Let R = the top k cases ranked according to sim
3. Return as $f(x)$ the class, c , that wins the majority vote among $f(R_1), f(R_2), \dots, f(R_{|k|})$. Handle ties randomly.

Slide CS472 – Machine Learning 12

Types of Attributes

1. Symbolic (nominal) – $EyeColor \in \{brown, blue, green\}$
2. Boolean – $anemic \in \{TRUE, FALSE\}$
3. Numeric (Integer, Real) – $age \in [0, 105]$

How do we compute the similarity between
 $EyeColor = brown$ and $EyeColor = green$?

Slide CS472 – Machine Learning 13

Example of case retrieval for k-nn

outlook	temp	humidity	windy	plan	<i>sim</i>
sunny	hot	high	false	cs472	
sunny	hot	high	true	cs472	
overcast	hot	high	false	soccer	
overcast	mild	normal	true	football	
rain	mild	high	false	soccer	
rain	cool	normal	false	soccer	

A : outlook, temp, humidity, windy

$k = 1$, $C = \{\text{soccer, cs472 football}\}$

test case: $X = \text{sunny cool high false}$

Slide CS472 – Machine Learning 14

k -Nearest Neighbor Algorithm

1. Memorizes all observed instances and their class
2. Is this rote learning?
3. Is this really learning?
4. When does the induction take place?

Slide CS472 – Machine Learning 15

Advantages and Disadvantages

What constitutes the concept description?

Slide CS472 – Machine Learning 16