CLASSIFICATION METHODS

Nearest-Neighbor classifiers

instance-based learning

eager learners

(1) constructing a classification model from data
(2) applying the model to test examples

uses specific training instances to make predictions without having to maintain an abstraction (or model) derived from data

Lazy learners

Set of Stored Cases

Atr1	………	AtrN	Class
			A
			B
			B
			C
			A
			C
			B

Unseen Case

Atr1	………	AtrN

Nearest-Neighbor classifiers

Rote-learner

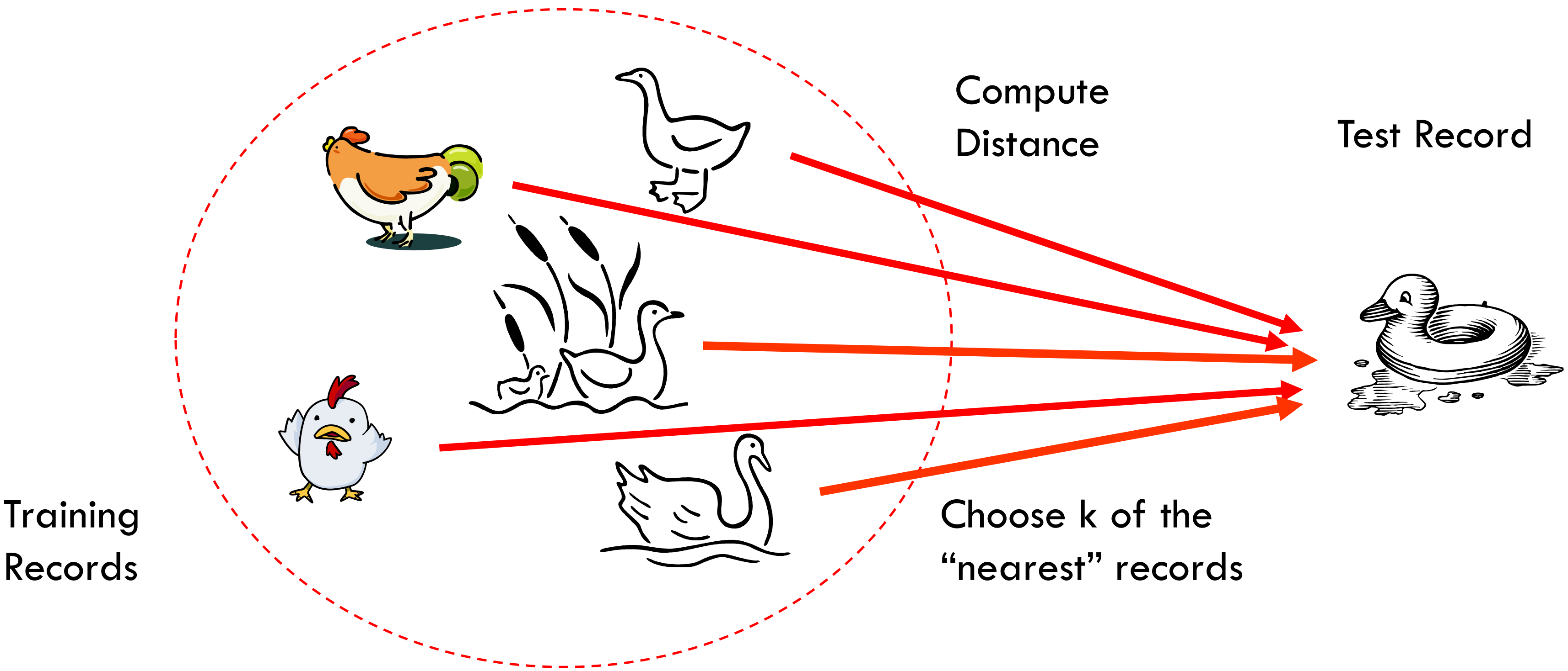
- Memorizes entire training data and performs classification only if attributes of record match one of the training examples exactly

Nearest neighbor

- Uses k “closest” points (nearest neighbors) for performing classification

Nearest Neighbor Classifiers

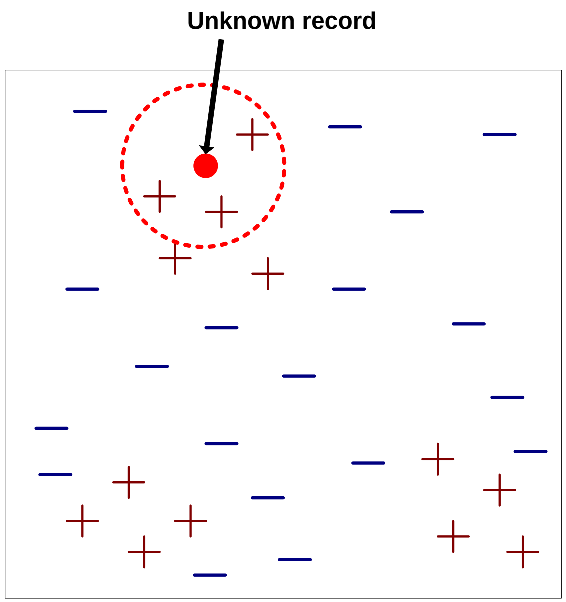
- Basic idea:
  - If it walks like a duck, quacks like a duck, then it's probably a duck



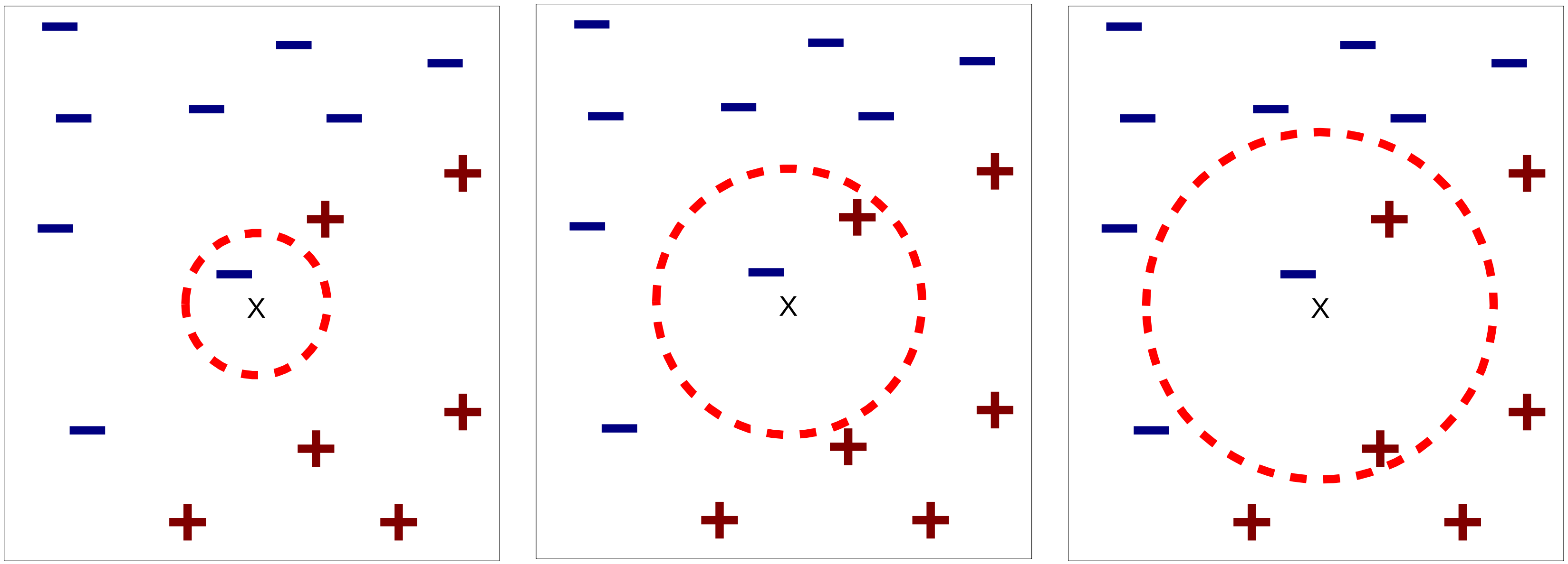
Nearest-Neighbor Classifiers

- Requires three things
  - The set of stored records
  - Distance Metric to compute distance between records
  - The value of *k*, the number of nearest neighbors to retrieve
- To classify an unknown record:
  - Compute distance to other training records
  - Identify *k* nearest neighbors
  - Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)

$$\text{Majority Voting:} \quad y' = \underset{v}{\operatorname{argmax}} \sum_{(\mathbf{x}_i, y_i) \in D_z} I(v = y_i)$$



Definition of Nearest Neighbor



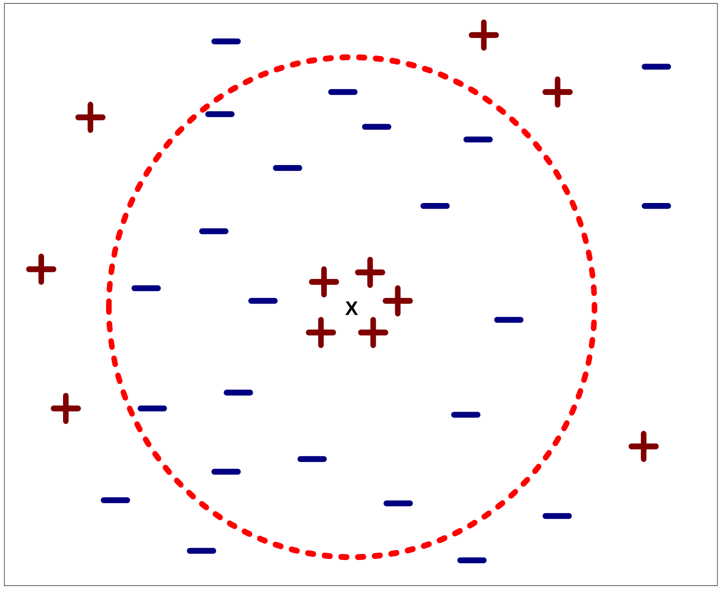
(a) 1-nearest neighbor (b) 2-nearest neighbor (c) 3-nearest neighbor

K-nearest neighbors of a record x are data points that have the k smallest distance to x

Nearest-Neighbor Classifiers

Choosing the value of k:

- ❖ If k is too small, sensitive to noise points
- ❖ If k is too large, neighborhood may include points from other classes



Nearest-Neighbor Classifiers

Distance-Weighted Voting: $y' = \underset{v}{\operatorname{argmax}} \sum_{(\mathbf{x}_i, y_i) \in D_z} w_i \times I(v = y_i)$

$$\boxed{w_i = 1/d(\mathbf{x}', \mathbf{x}_i)^2}$$

- ✓ Lazy learners
- ✓ Nearest-neighbor classifiers can produce arbitrarily shaped decision boundaries
- ✓ appropriate proximity measure
- ✓ data preprocessing steps

Bayes Classifier

- A probabilistic framework for solving classification problems
- Conditional Probability:

$$P(C \mid A) = \frac{P(A, C)}{P(A)}$$

$$P(A \mid C) = \frac{P(A, C)}{P(C)}$$

- Bayes theorem:

$$P(C \mid A) = \frac{P(A \mid C) P(C)}{P(A)}$$

Example of Bayes Theorem

- Given:
  - A doctor knows that meningitis causes stiff neck 50% of the time
  - Prior probability of any patient having meningitis is 1/50,000
  - Prior probability of any patient having stiff neck is 1/20

- If a patient has stiff neck, what's the probability he/she has meningitis?

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

Bayesian Classifiers

- Consider each attribute and class label as random variables

- Given a record with attributes $(x_1, x_2,\ldots,x_n)$
  - Goal is to predict class Y
  - Specifically, we want to find the value of Y that maximizes $P(Y \mid x_1, x_2,\ldots,x_n)$

- Can we estimate $P(Y \mid x_1, x_2,\ldots,x_n)$ directly from data?

Bayesian Classifiers

- Approach:
  - compute the posterior probability $P(Y \mid x_1, x_2, \ldots, x_n)$ for all values of Y using the Bayes theorem

$$\boxed{P(Y \mid x_1 x_2 \ldots x_n) = \frac{P(x_1 x_2 \ldots x_n \mid Y) P(Y)}{P(x_1 x_2 \ldots x_n)}}$$

  - Choose value of Y that maximizes
    $P(Y \mid x_1, x_2, \ldots, x_n)$

  - Equivalent to choosing value of Y that maximizes
    $P(x_1, x_2, \ldots, x_n \mid Y) P(Y)$

- How to estimate $P(x_1, x_2, \ldots, x_n \mid Y)$?

Naïve Bayes Classifier

- Assume independence among attributes A_i when class is given:
  - $P(x_1, x_2, \ldots, x_n \mid Y) = P(x_1 \mid Y_j)\, P(x_2 \mid Y_j) \ldots P(x_n \mid Y_j)$
  - Can estimate $P(x_i \mid Y_j)$ for all x_i and Y_j.
  - New point is classified to Y_j if $P(Y_j) \prod P(x_i \mid Y_j)$ is maximal.

How to Estimate Probabilities from Data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Class: $P(Y_k) = N_k/N$
  - e.g., $P(\text{No}) = 7/10$,
    $P(\text{Yes}) = 3/10$
- For discrete attributes:

$$P(x_i \mid Y_k) = |x_{ik}| / N_k$$

  - where $|x_{ik}|$ is number of instances having attribute x_i and belongs to class Y_k
  - Examples:

$P(\text{Status=Married}|\text{No}) = 4/7$
$P(\text{Refund=Yes}|\text{Yes})=0$

How to Estimate Probabilities from Data?

- For continuous attributes:
  - Discretize the range into bins
  - Probability density estimation:
    - Assume attribute follows a normal distribution
    - Use data to estimate parameters of distribution (e.g., mean and standard deviation)
    - Once probability distribution is known, can use it to estimate the conditional probability $P(x_i|c)$

How to Estimate Probabilities from Data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Normal distribution:

$$P(x_i \mid Y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

  - One for each (x_i, Y_i) pair

- For (Income, Class=No):
  - If Class=No
    - sample mean = 110
    - sample variance = 2975

$$P(Income = 120 \mid No) = \frac{1}{\sqrt{2\pi(54.54)}} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$