

INT247

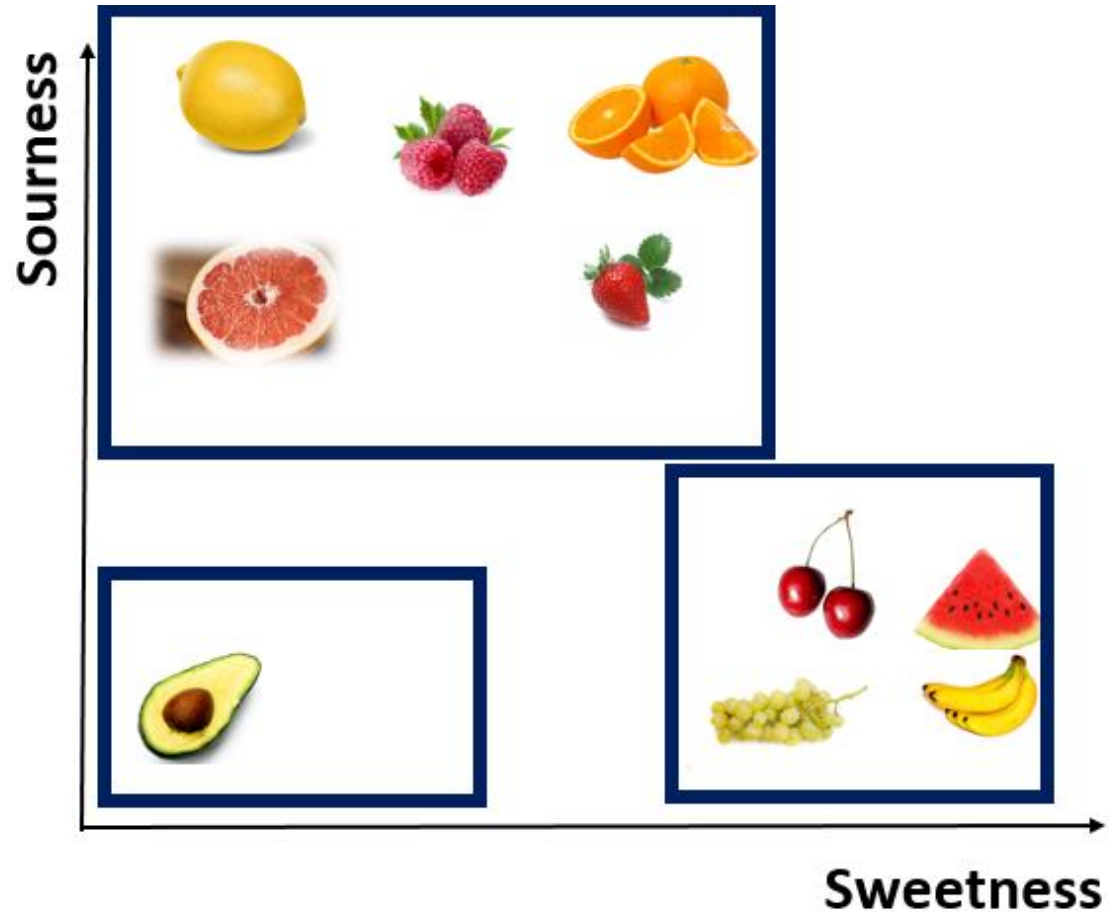
Machine Learning Foundations

Lecture #2.1

K-Nearest Neighbour and Bayesian Learning

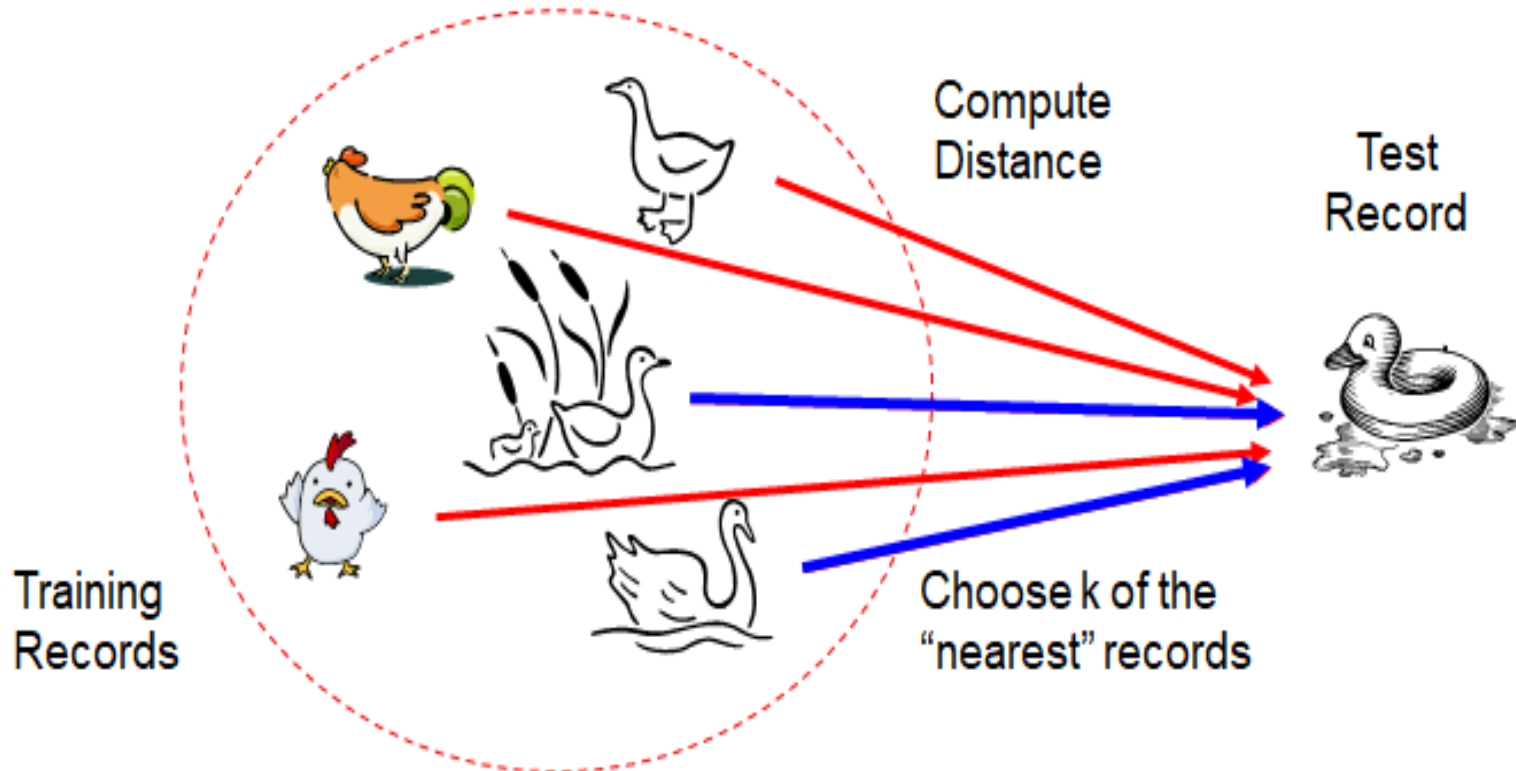
K- Nearest Neighbour Algorithm

- **Lazy Learner**
- **Uses K “closest” points for performing classification.**



K- Nearest Neighbour Algorithm

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



K- Nearest Neighbour Algorithm

- **Requires three things:**
 - **The set of stored records.**
 - **Distance Metric to compute distance between records.**
 - **The value of K, the number of nearest neighbours to retrieve .**
- **To classify an unknown record:**
 - **Compute distance to other training records.**
 - **Identify k nearest neighbours.**
 - **Use class labels of nearest neighbours to determine the class label of unknown record.**

Compute Distance

- Euclidean distance

$$d(p, q) = \sqrt{\sum_i (p_i - q_i)^2}$$

Numerical Example

- Consider a dataset with 2 attributes:

- Acid durability (x_1)

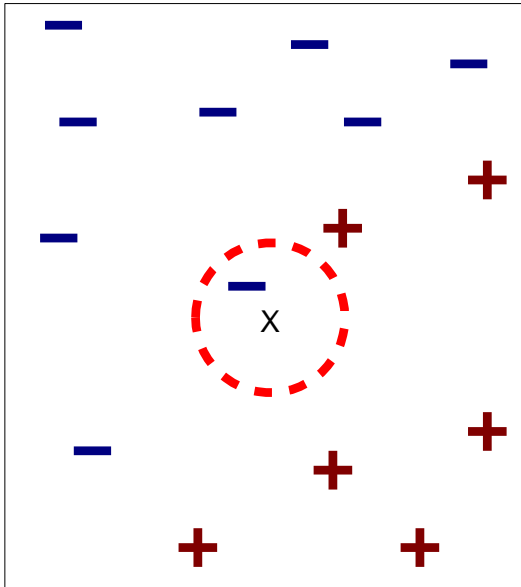
- Strength (x_2)

to classify whether a special paper tissue is good or bad (y).

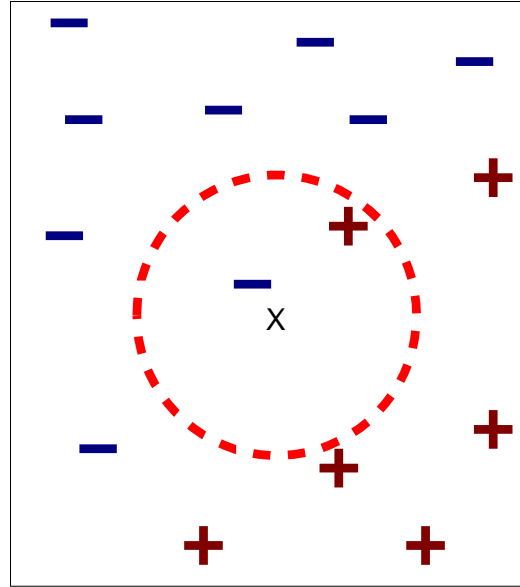
x_1	x_2	y
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

- Classify paper tissue with $x_1=3$ and $x_2=7$ considering $K=3$.

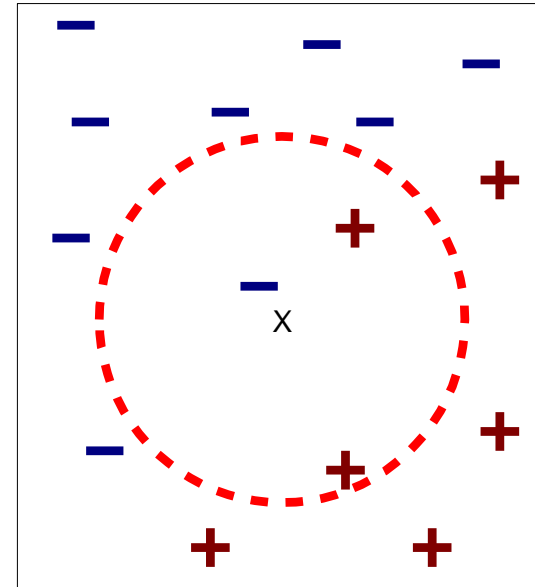
K- Nearest Neighbour Algorithm



(a) 1-nearest neighbor



(b) 2-nearest neighbor



(c) 3-nearest neighbor

Note:

- If K is too small, sensitive to noise points.
- If K is too large, neighbourhood may include points from other classes.

Scaling Issues

- **Attributes may have to be scaled to prevent distance measures from being dominated by one of the attributes.**
- **Examples:**
 - **Height of a person may vary from 1.5m to 1.8m**
 - **weight of a person may vary from 90lb to 300lb**
 - **income of a person may vary from \$10K to \$1M**

Advantages of KNN

- No assumption about data.
- Simple algorithm.
- High accuracy.
- Versatile – useful for classification or regression.

Disadvantages of KNN

- **Computationally expensive.**
- **High memory requirement.**
- **Store all the training data.**
- **Sensitive to irrelevant features and the scale of the data.**

Bayesian Learning

- **Probabilistic approach to inference.**
- **Quantities of interest are governed by probability distributions.**
- **Optimal decisions can be made by reasoning about these probabilities together with observed training data.**

Bayesian Learning

- Initial knowledge of many probabilities is required.
- Significant computational costs required.

Bayes Theorem

- Direct method of calculating the probability of best hypothesis.

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)}$$

- $P(h)$: prior probability of (h) hypothesis .
- $P(D)$: prior probability of (D) observation.
- $P(D|h)$: Probability of observing D in which h holds.
- $P(h|D)$: posterior probability of h, reflects confidence that h holds after D has been observed.

MAP (Maximum a Posteriori) Hypothesis

- $\mathbf{h}_{MAP} =$
$$\begin{aligned} h_{MAP} &= \arg \max_{h \in H} P(h|D) \\ &= \arg \max_{h \in H} \frac{P(D|h)P(h)}{P(D)} \\ &= \arg \max_{h \in H} P(D|h)P(h) \end{aligned}$$

- **P(D) can be dropped, because it is a constant independent of h.**

Example

- Consider a medical diagnosis problem where:
 - $P(\text{cancer})=0.008$
 - $P(\text{pos} | \text{cancer})=0.98$
 - $P(\text{pos} | \sim\text{cancer})=0.03$
- If a new patient comes in with a positive test result, what is the probability that he has cancer?
$$P(\text{pos} | \text{cancer}).P(\text{cancer})=0.98*0.008=0.0078$$
$$P(\text{pos} | \sim\text{cancer}).P(\sim\text{cancer})=0.03*0.992=0.0298$$

Thus, $h_{\text{MAP}} = \sim\text{cancer}$

Brute Force MAP Learning

- In order to specify a learning problem of the algorithm, values for $P(h)$ and $P(D|h)$ must be specified.
- Assumptions:
 - Training data D is noise free (i.e. $d_i = c(x_i)$)
 - Target concept c is contained in H i.e.
$$(\exists h \in H)[(\forall x \in X)[h(x) = c(x)]]$$
 - No reason to believe that any hypothesis is more probable than any other.

Learning a Real-Valued function

- Consider any real-value target function f .
- Training examples $\langle x_i, d_i \rangle$, where d_i is noisy training value.
- $d_i = f(x_i) + e_i$
- Where e_i is random variable drawn independently for each x_i according to some Gaussian distribution with mean=0.
- The maximum likelihood hypothesis h_{ML} is defined as:

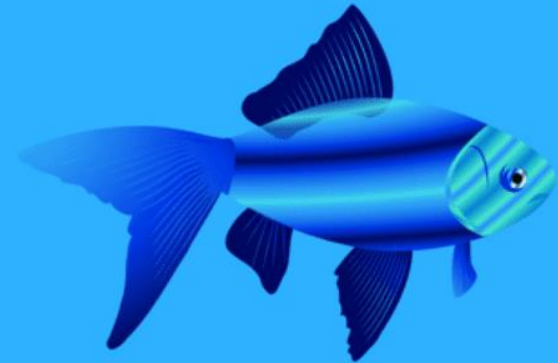
$$h_{ML} = \operatorname{argmax}_{h \in H} p(D | h)$$

$$= \operatorname{argmax}_{h \in H} \prod_{i=1}^m p(d_i | h)$$

$$h_{ML} = \operatorname{argmin}_{h \in H} \sum_{i=1}^m (d_i - h(x_i))^2$$

Example of Naïve Bayes Classifier

- **Dataset with 3 classes:**
 - Parrot
 - Dog
 - Fish
- **4 features:**
 - Swim
 - Wings
 - Green Color
 - Dangerous Teeth



@dataaspirant.com

Example of Naïve Bayes Classifier

Swim	Wings	Green Color	Dangerous Teeth	Animal Type
50	500/500	400/500	0	Parrot
450/500	0	0	500/500	Dog
500/500	0	100/500	50/500	Fish

- **Table shows a frequency table of dataset.**

Example of Naïve Bayes Classifier

- **Considering the dataset predict the type of animal if:**
 - **Swim=True**
 - **Wings=False**
 - **Green=True**
 - **Teeth=False**

Example of Naïve Bayes Classifier

- Use naïve bayes approach for multiple evidences:

$$\begin{aligned} &P(H|E_1, E_2 \dots E_N) \\ &= \frac{P(E_1|H) * P(E_2|H) \dots * P(E_N|H) * P(H)}{P(E_1, E_2, \dots E_N)} \end{aligned}$$

Example of Naïve Bayes Classifier

- For hypothesis testing for animal to be a dog:

$$\begin{aligned} P(\text{Dog} | \text{Swim}, \\ \text{Green}) &= P(\text{Swim} | \text{Dog}) * P(\text{Green} | \text{Dog}) * P(\text{Dog}) / P(\text{Swim}, \text{Green}) \\ &= 0.9 * 0 * 0.333 / P(\text{Swim}, \text{Green}) \\ &= 0 \end{aligned}$$

Example of Naïve Bayes Classifier

- For hypothesis testing for animal to be a Parrot:

$$\begin{aligned} P(\text{Parrot} | \text{Swim}, \\ \text{Green}) &= P(\text{Swim} | \text{Parrot}) * P(\text{Green} | \text{Parrot}) * P(\text{Parrot}) / \\ &P(\text{Swim}, \text{Green}) \\ &= 0.1 * 0.8 * 0.333 / P(\text{Swim}, \text{Green}) \\ &= 0.0264 / P(\text{Swim}, \text{Green}) \end{aligned}$$

Example of Naïve Bayes Classifier

- For hypothesis testing for animal to be a Fish:

$$\begin{aligned} P(\text{Fish} | \text{Swim}, \\ \text{Green}) &= P(\text{Swim} | \text{Fish}) * P(\text{Green} | \text{Fish}) * P(\text{Fish}) / P(\text{Swim}, \text{Green}) \\ &= 1 * 0.2 * 0.333 / P(\text{Swim}, \text{Green}) \\ &= 0.0666 / P(\text{Swim}, \text{Green}) \end{aligned}$$

Example of Naïve Bayes Classifier

- For all the three cases, denominator is same.
- The value of $P(\text{Fish} | \text{Swim, Green})$ is greater than $P(\text{Parrot} | \text{Swim, Green})$.
- Thus, Class of animal is **Fish**.

Exercise

- **Predict the class of animal if:**
 - **Swim=True**
 - **Wings=False**
 - **Green=True**
 - **Teeth=True**



COMING UP
