greatlearning
*Learning for Life*


# MACHINE LEARNING- WEEK-1

greatlearning
Learning for Life

# DSBA CURRICULUM DESIGN

## FOUNDATIONS

- Data Science Using Python
- Statistical Methods for Decision Making

## CORE COURSES

- Advanced Statistics
- Data Mining
- Predictive Modelling
- Machine Learning(Week-1/5)
- Time Series Forecasting
- Data Visualization
- SQL

## DOMAIN APPLICATIONS

- Financial Risk Analytics
- Marketing Retail Analytics

greatlearning
*Learning for Life*

# LEARNING OBJECTIVE OF THIS MODULE

- Supervised Learning : KNN & Naïve Bayes
- Ensemble Techniques: Bagging, Boosting, Cross-validation and SMOTE
- Text Mining & Sentiment Analysis

greatlearning
Learning for Life

# LEARNING OBJECTIVES OF THIS SESSION

- Naïve Bayes
- K Nearest Neighbour (KNN)

greatlearning
*Learning for Life*

# TRY ANSWERING THE FOLLOWING

- Is KNN a Non-Parametric Method?
- Does Naive Bayes work on conditional probability?
- Naïve Bayes is “naïve” because of its assumptions?

greatlearning
Learning for Life

# BROAD OVERVIEW – Naïve Bayes

What we want to know; the Posterior probability of Class j given a predictor x

The Likelihood; the probability of the predictor given a Class j. Its computed from the training-set.

The Prior probability of Class j; what we know about the class distribution before we consider x.

$$P(Class_j \mid x) = \frac{P(x \mid Class_j) \times P(Class_j)}{P(x)}$$

The Evidence. In practice, there's interest only in the numerator (denominator is effectively constant)

Applying the independence assumption

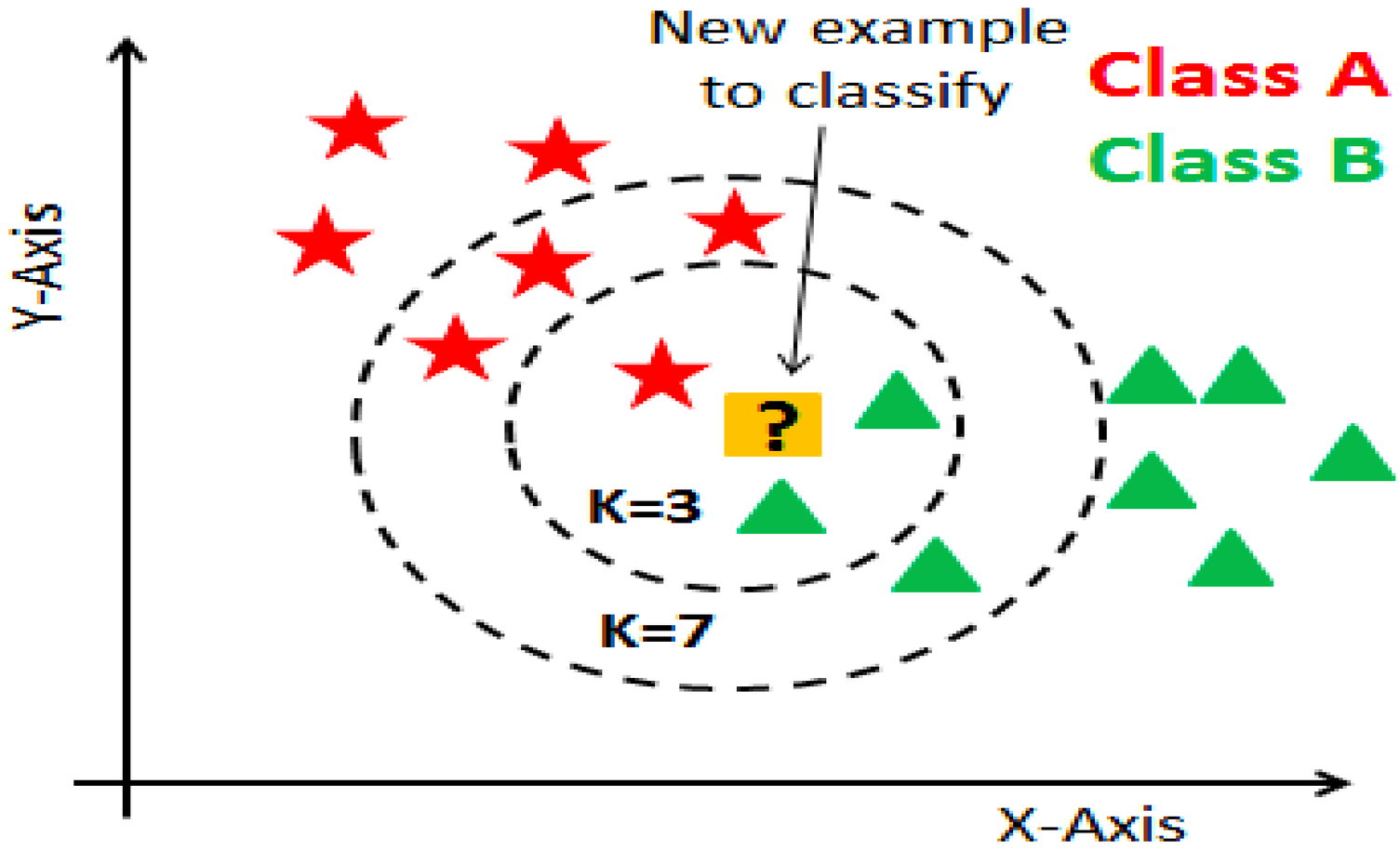
$$P(x \mid Class_j) = P(x_1 \mid Class_j) \times P(x_2 \mid Class_j) \times \ldots \times P(x_k \mid Class_j)$$

Substituting the independence assumption, we derive the Posterior probability of Class j given a new instance x' as...

$$P(Class_j \mid x') = P(x'_1 \mid Class_j) \times P(x'_2 \mid Class_j) \times \ldots \times P(x'_k \mid Class_j) \times P(Class_j)$$

greatlearning
Learning for Life

# BROAD OVERVIEW - KNN



greatlearning
Learning for Life

# Industry Application

The modern systems are now able to use k-nearest neighbor for visual pattern recognition to scan and detect hidden packages in the bottom bin of a shopping cart at check-out. If an object is detected that's an exact match for an object listed in the database, then the price of the spotted product could even automatically be added to the customer's bill. While this automated billing practice is not used extensively at this time, the technology has been developed and is available for use.

K-nearest neighbor is also used in retail to detect patterns in credit card usage. Many new transaction-scrutinizing software applications use kNN algorithms to analyze register data and spot unusual patterns that indicate suspicious activity.

greatlearning
Learning for Life

# CASE STUDY- Project Choice

University of Torinto is to launch some new programs and want to know about there target audience(that is Which student should be pitched for which course/program). They hired you to do the analysis and predict which program a student will choose out of Academic Programs and Vocational Programs on the basis of the information given such as socio-economic status, the type of school attended (public or private), gender and their prior reading, writing, math and science scores.

# Data Science @ Work

greatlearning
Power Ahead

Apply **Data Science at your workplace** to gain some instant benefits:

- Get noticed by your management with your outstanding analysis backed by data science.
- Create an impact in your organization by taking up small projects/initiatives to solve critical issues using data science.
- Network with members from the data science vertical of your organization and seek opportunities to contribute in small projects.
- Share your success stories with us and the world to position yourself as a subject matter expert in data science.

greatlearning
*Learning for Life*

# ANY QUESTIONS

greatlearning
*Learning for Life*

# HAPPY LEARNING