

MACHINE LEARNING- WEEK-1

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**

LEARNING OBJECTIVE OF THIS MODULE

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

LEARNING OBJECTIVES OF THIS SESSION

- Naïve Bayes
- K Nearest Neighbour (KNN)

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?



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 .

$$P(\text{Class}_j | x) = \frac{P(x | \text{Class}_j) \times P(\text{Class}_j)}{P(x)}$$

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

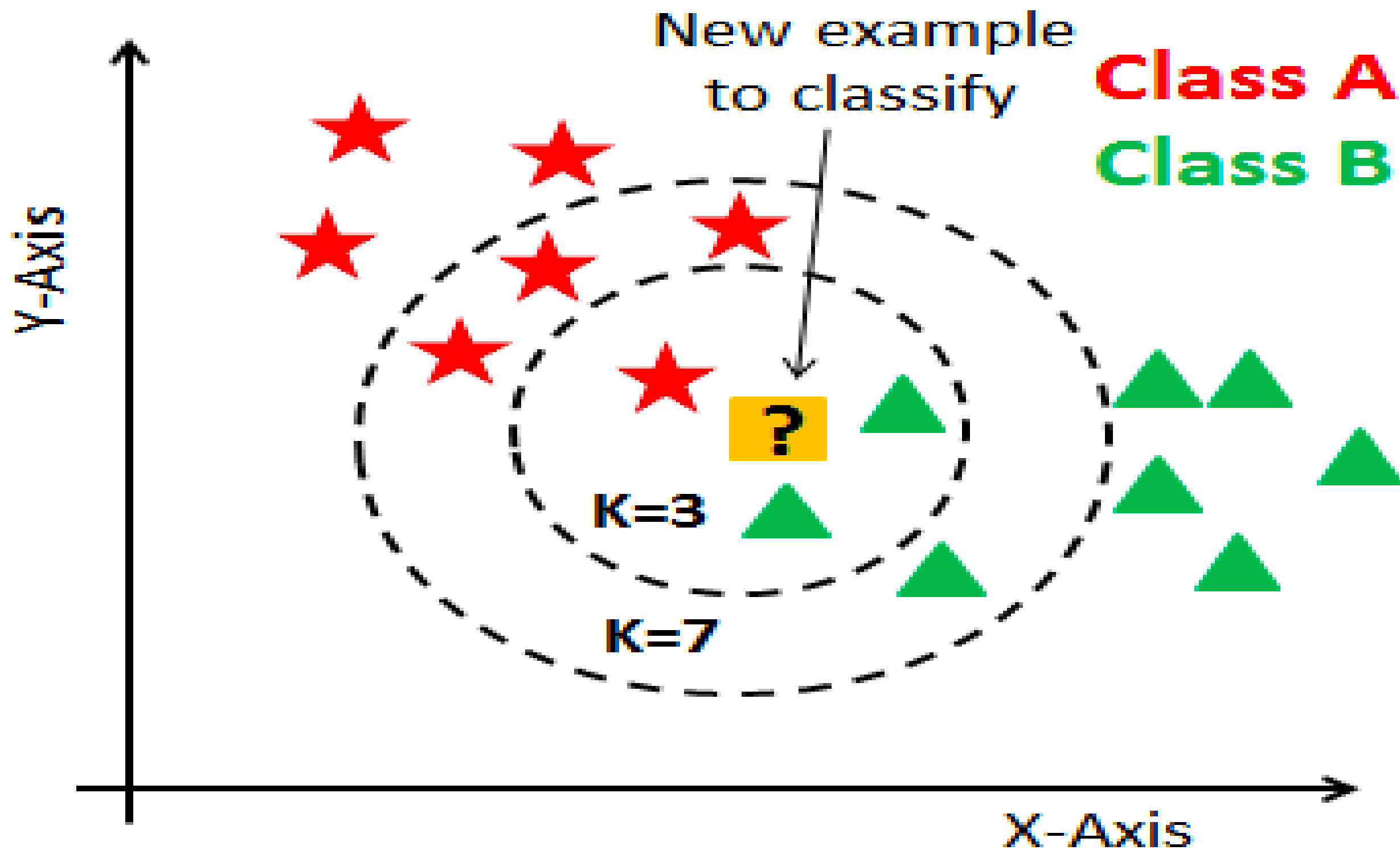
Applying the **independence** assumption

$$P(x | \text{Class}_j) = P(x_1 | \text{Class}_j) \times P(x_2 | \text{Class}_j) \times \dots \times P(x_k | \text{Class}_j)$$

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

$$P(\text{Class}_j | x') = P(x'_1 | \text{Class}_j) \times P(x'_2 | \text{Class}_j) \times \dots \times P(x'_k | \text{Class}_j) \times P(\text{Class}_j)$$

BROAD OVERVIEW - KNN



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.

CASE STUDY- Project Choice

University of Toronto 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

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.



ANY QUESTIONS



HAPPY LEARNING

Excelerate CareerPrep:

Career services by Great Learning.

The CareerPrep is an organized way to learn about all of the topics asked in interviews for Data Science jobs.

- Complete Excelerate career prep and learn how to crack an interview with ease.
- Understand the different kind of roles available in the data science ecosystem along with their skill set requirements
- Build a resume that has a higher probability of getting shortlisted by recruiters
- Prepare for different data science roles by reviewing 400+ questions asked in actual interviews
- Understand and implement frameworks to solve guesstimates, logical and aptitude problems often used in interviews
- Learn from the experience of alumni who have successfully transitioned
- Speed up your path towards a transition to a Data science role.

Complete Career Prep today to speed up your path towards DS transition. You'll have a 40% better chance of getting hired than other candidates.

Note: CareerPrep is mandatory step in unlocking the job board.