Lesson 14

Naïve-Bayes and Support Vector Machines (SVMs) Classifiers

Naïve-Bayes Classifier

- Parallel algorithm
- Widely Preferred Text Analytics
- Medium to Large Datasets 1M to 100M training examples which take too long time on SGD (Sequential, online incremental execution) or SVM (Sequential execution)

Naïve Bayes Classifier

- Naïve means unsophisticated, ..., a simple classifier
- Probabilistic and statistical classifier
- Based on Bayes theorem (from Bayesian statistics) with assumption of strong (Naïve) independence and maximum posteriori (MAP) hypothesis

Naïve Bayes Classifier

- A supervised learning technique, which uses non-parametric approach
- Uses assumption that features have strong independences
- "maximum a posteriori (MAP)" used to obtain the most likely class (Posteriori means at the back of something. For example, hypothesis)

Bayes Classification Assumption

- Naïve independence assumptions (conditional independence)
- The classifier computes condition probabilities for the conditional independence

Document classification in Text Analytics

- Use the bag-of-words model
- The pre-processing of a document first provides a document with a bag of words
- The occurrence (frequency) of each word as a feature used for training a classifier
- [Refer Section 9.2.2 Example 9.3]

Bayes Classification

- Probability that a bag-of-words x belong to k^{th} class equals the product of individual probabilities of those words.

$$P(\mathbf{x}|c_k) = \prod_{i=1}^{n} P(x_i|c_k),$$ where x_i is a discrete random variable (word), $i = 1, 2, \ldots, n$, when n is number of words in the bag.

Meaning of Symbols

- Π is sign for the product of n terms. P(xi|ck) means probability of condition that state the value = x_i and of c = c_k

Training Data: (Car-Model, Annual Income, Age)

- Car Models: Jaguar XJ (JXJ), Jaguar XF (JXF), Harrier (H), Zest (Z)
- Cost in Million of Rupees: JXJ = 10.1, JXF = 4.977, H = 1.269, and Zest (Z) = 0.554
- Price and AI are car price and annual income, both in Million of Rupees (Rs.)

Buyers Training Data: (Car-Model, AI, Age)

- (JXJ, 142, 39), (JXF, 80, 49), (H, 48, 43), (Z, 28, 46), (JXJ, 138, 44), (JXF, 82, 45), (H, 52, 40), (Z, 24, 34),
- (JXJ, 140, 44), (JXF, 70, 36), (H, 58, 38), (Z, 28, 46), (JXJ, 162, 36), (JXF, 86, 52), (H, 43, 33), (Z, 23, 36), (JXJ, 132, 44), (JXF, 90, 46), (H, 48, 42), (Z, 18, 26),...

Computations of Means and Variances

Car Model	Mean Annual Income (AI) in MRs.	σ_{AI} MRs.	Age Years	σ_{Age} Years
JXJ	140	40	40	8
JXF	80	30	48	16
H	50	20	42	8
Z	25	15	40	20

Probabilities

- P(JXJ) = Probability of Buying model JXJ among four models
- P(JXF), P(H) and P(Z) are probabilities of buying models JXF, H and Z, respectively

Conditional independence

- Assume that feature probabilities P (Car Model): P(JXJ), P(JXF), P(H) and P(Z) are independent given the class of annual income and age of the buyers

Conditional Probabilities

- Conditional Probabilities p(AI|JXJ) and p(Age|JXJ) are probabilities that JXJ buyer income is AI and buyer age is Age, respectively
- p(AI|JXF) and p(Age|JXF), p(AI|H) and p(Age|H), and p(AI|Z) and p(Age|Z) are conditional probabilities for buying models JXF, H and Z, respectively

Computations of Posterior

- Posterior (JXJ) = $E^{-1}\{P(JXJ).p(AI|JXJ).p(Age|JXJ)$
- Posterior (JXF) = $E^{-1}\{P(JXF).p(AI|JXF).p(Age|JXF)$
- Posterior (H) = $E^{-1}\{P(H).p(AI|H).p(Age|H)$
- Posterior (JXJ) = $E^{-1}\{P(Z).p(AI|Z).p(Age|Z)$

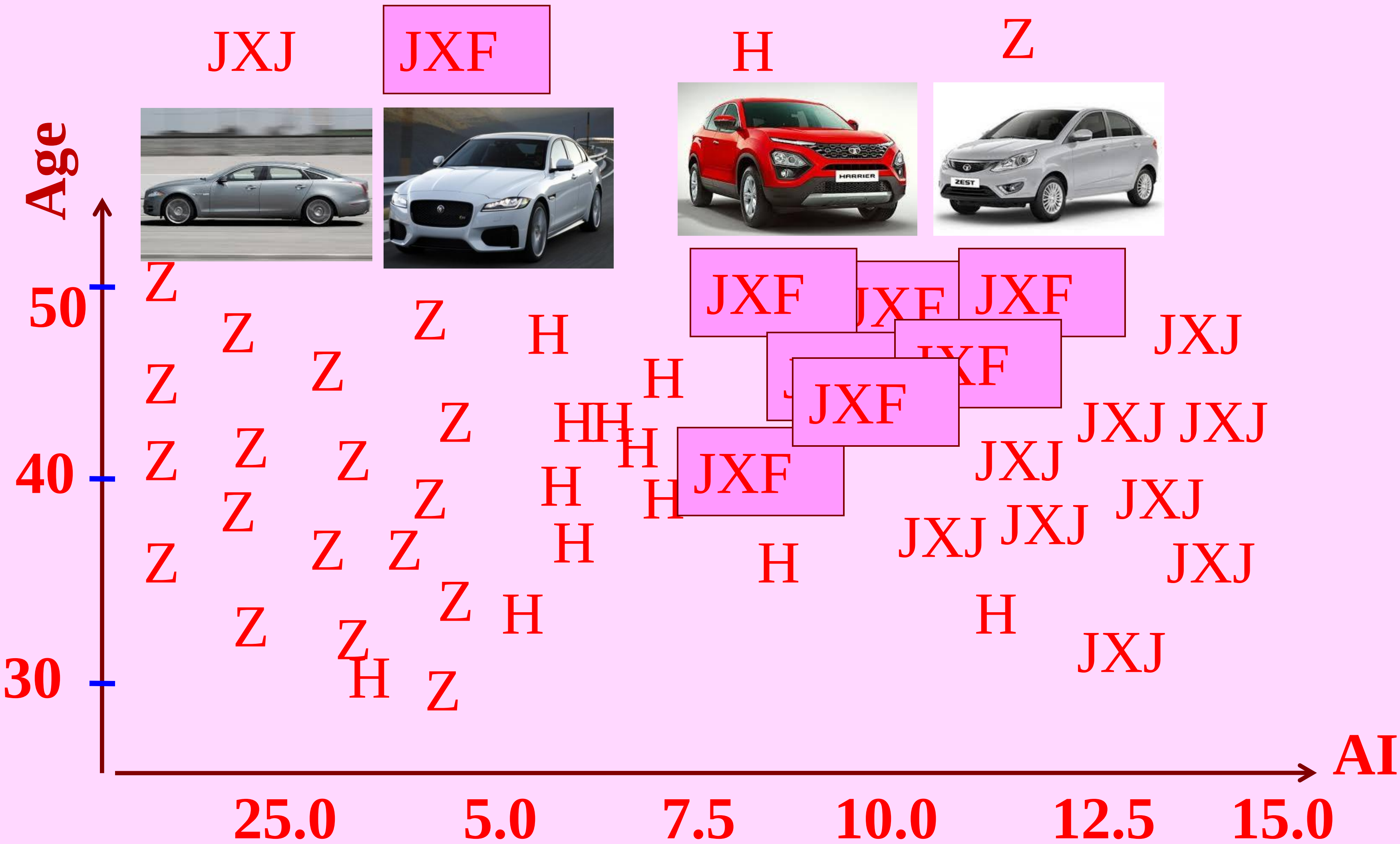
Maximum a Posteriori (MAP) (to obtain most likely class) decision rule

- The arguments of the maxima, argmax are the points of some function at which the function values (outputs)are maximized refers to the *inputs*
- $C_{MAP} = \underset{c \in C}{\text{argmax}} \{\text{Posteriori (JXJ)}\}$. c is
- a class with members (Car model, AI, Age)

Computation of Evidence

- $E = \{P(JXJ).p(AI|JXJ).p(Age|JXJ) +$
- $\{P(JXF).p(AI|JXF).p(Age|JXF) +$
- $\{P(H).p(AI|H).p(Age|H) +$
- $\{P(Z).p(AI|Z).p(Age|Z)$

Classification of Car Models with AI in Million Rupees and Buyers Age and



Applications

- Pattern recognition,
- image analysis,
- information retrieval
- bioinformatics

Support Vector Machine (SVM) Classifier

- Sequential, Sleek and Efficient in appropriate data range
- A method in a set of related supervised learning method that uses a vector, which has in general, v elements in v-dimensional space
- The vector classifies the data points.

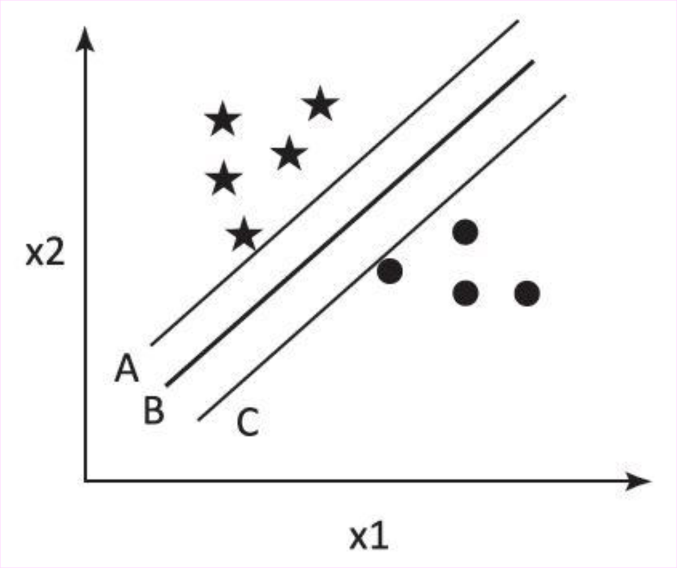
Support Vector Machine (SVM) Classifier

A data point in the space is represented by a vector. A data point represents by (x1, x2, ..., xn) in n-dimensional space. Consider two-dimensional space, with data points (x1, x2) and axes X1 and X2. Each data-point if considered as a vector element has two components, x1, and x2. (Two sets of words in text analysis

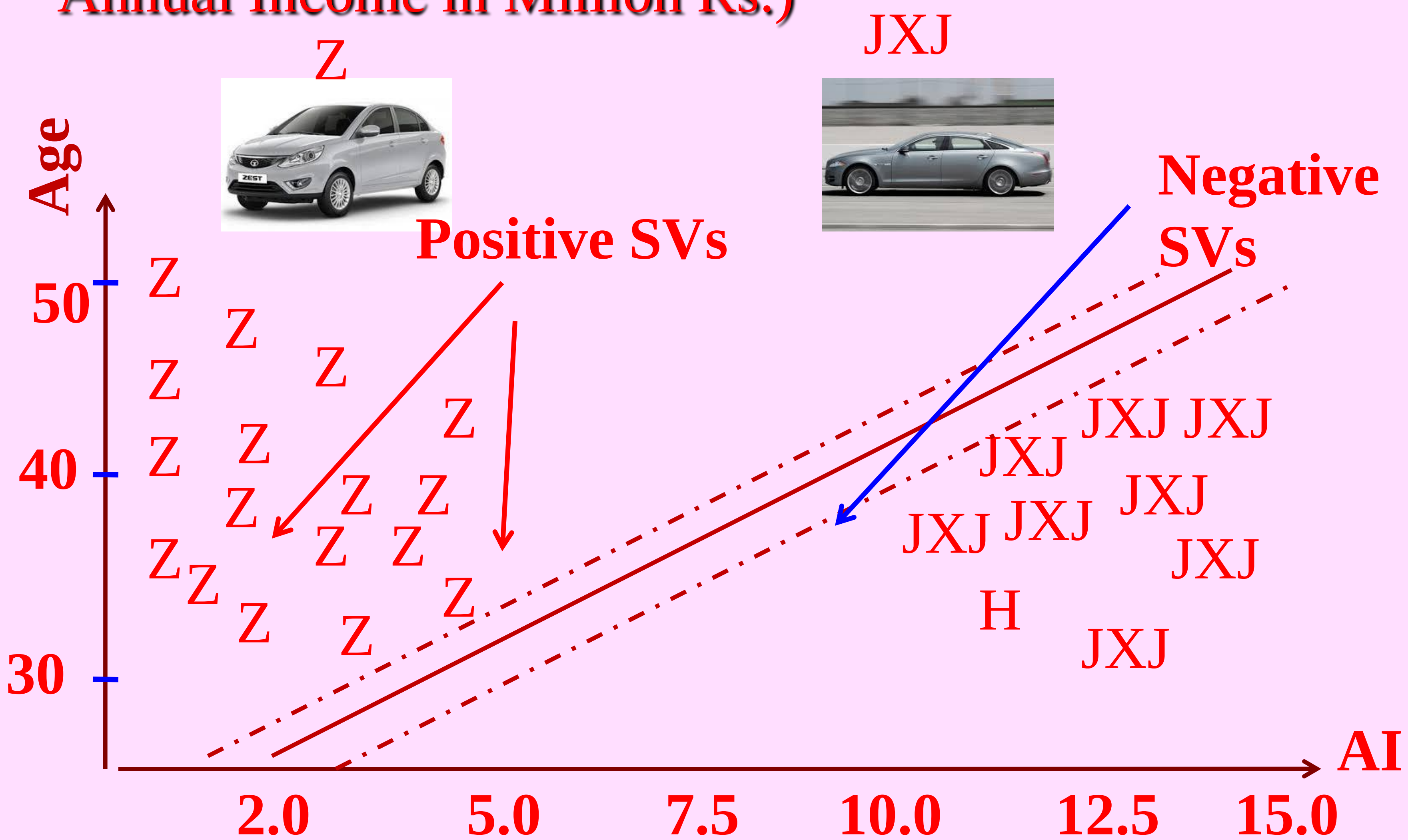
Hyperplane

- A subspace of one dimension less than its ambient space in geometry
- If a space is 3-dimensional, then its hyperplanes are the 2-dimensional planes
- However, if the space is 2-dimensional, its hyperplanes are 1-dimensional which means lines.

Figure 6.18 Three hyperplanes A, B, and C for classification of data points



Classification of Zest (Z) and Jaguar XJ (JXJ) car models Using support vector hyperplanes (AI = Annual Income in Million Rs.)



Data Points in the Figure

- Data points for two car models shown in figure
- Solid line Hyperplane and two dashed line hyperplanes are shown.
- Training data-points fit a linear kernel such that deviations

Hyperplanes and Boundary planes

- The planes are iteratively chosen to maximize distances. Positive and negative support vectors can be used in features space.
- Negative SV means the features, which exclude in classification, are used by the classifier.

Training Algorithm

- Generates boundary plane defining vectors such that all or most data points are at distance from the boundary
- Mid-points of the boundaries generates 'Support Vector'
- Such that maximize boundary distances
- An objective function created such that the boundaries are at maximum distance

Nonlinear kernels

- For example, Quadratic Kernel Function: Circle, parabola, ellipse, ellipsoid equations
- Objective functions and classifier vectors generated from the quadratic function Applications: Support Vectors

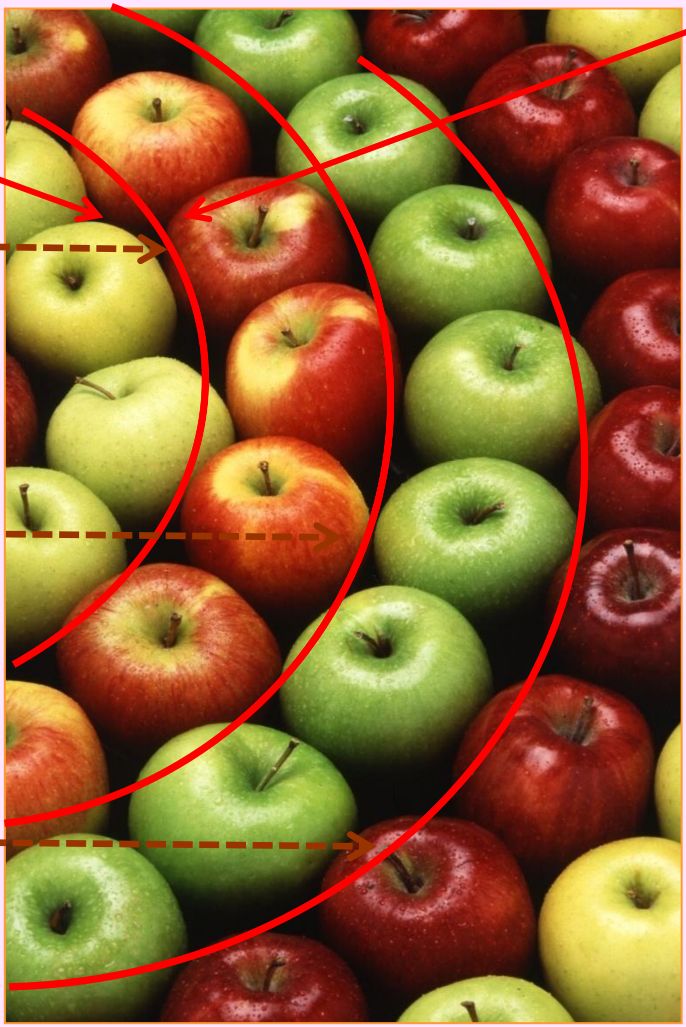
Three Quadratic Kernel Functions for classifying

Positive SVs

Quadratic function K1

Quadratic function K2

Quadratic function K3



Negative SVs for yellow-green apples

Four Classes of Apples using three Support Vectors based on Quadratic Kernel

Summary

We learnt:

- Naïve-Bayes Classifier
- Support Vector Machines (SVMs) Classifiers

End of Lesson 14 on

Naïve-Bayes and Support Vector Machines (SVMs) Classifiers