# INTRODUCTION TO STATISTICAL LEARNING THEORY

* **FEATURE**:- The feature is an attribute of a data set that is used in a ML process.

→ The features in a data set are also called its dimensions. So a data set having 'n' features is called an n-dimensional data set.

Eg **IRIS DATA SET**:- It has 5 attributes or features namely Sepal Length, Sepal Width, Petal length, Petal Width & Species
① ② ③ ④ ⑤
(① ② ③ ④ → Predictor Variables; ⑤ → Class Variable)

* (5-Dimensional Data Set)

* **FEATURE ENGINEERING**:- It refers to process of translating a data set into features such that these features are able to represent the data set more efficiently and result in a better learning performance.

① **Feature Transformation** → Transforms the data structured or unstructured into a new set of features which can represent the underlying problem which ML is trying to solve.

→ **Feature Construction**:- Discovers missing information about the relationships b/w features & augments the features space by creating additional features.

→ **Feature Extraction**:- Process of extracting or creating new set of features from the original set of features using some functional mapping.

② **Feature Selection**

→ It derives a subset of features from the full feature set which is most meaningful in the context of a specific ML problem

## Feature Transformation :- (FT)

↔ Engg. a good feature space is a crucial pre-requisite for the success of any ML model.

→ FT is used as an effective tool for dimensionality reduction & hence for boosting ML model performance.

→ Broadly there are 2 distinct goals of FT:-

a) Achieving best reconstruction of the original features in the data set.

b) Achieving highest efficiency in the learning task.

i) Feature Construction (FC) :- It involves transforming a given set of i/p features to generate a new set of more powerful features. For following situations, FC is required.

ii) when features have categorical value & ML needs numeric value i/ps.

ii) when features having numeric (cont.) values & need to converted to ordinal values.

iii) when text-specific FC needs to be done.

Eg REAL-ESTATE DATA SET :-

| | FLAT length | FLAT Breadth | FLAT PRICE |
|---|---|---|---|
| ① | 80 | 59 | 23,60,000 |
| ② | 54 | 45 | 12,15,000 |
| ③ | 78 | 56 | 21,84,000 |
| ④ | 63 | 63 | 19,84,000 |
| ⑤ | 83 | 74 | 30,71,000 |
| ⑥ | 92 | 86 | 39,56,000 |

Table-1

| | FLAT length | FLAT Breadth | FLAT Area | FLAT Price |
|---|---|---|---|---|
| ① | 80 | 59 | 4,720 | 23,60,000 |
| ② | 54 | 45 | 2,430 | 12,15,000 |
| ③ | 78 | 56 | 4,368 | 21,84,000 |
| ④ | 63 | 63 | 3,969 | 19,84,000 |
| ⑤ | 83 | 74 | 6,142 | 30,71,000 |
| ⑥ | 92 | 86 | 7,912 | 39,56,000 |

Table-2

## FC (a) Encoding Categorical (nominal) variables :-

| Age (Years) | City of origin | Parents athlete | Chance of win |
|---|---|---|---|
| 18 | City A | Yes | Y |
| 20 | City B | No | Y |
| 23 | City B | Yes | Y |
| 19 | City A | No | N |
| 18 | City C | Yes | N |
| 22 | City B | Yes | Y |

(a)

| Age (Years) | origin_ city_A | origin_ city_B | origin_ city_C | parents_ athlete_Y | parents_ athlete_N | win_ chance_Y | win_ chance_N |
|---|---|---|---|---|---|---|---|
| 18 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| 20 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 23 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 19 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| 18 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| 22 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |

(b)

| Age (Years) | origin_city_A | origin_city_B | origin_city_C | parents_athlete_Y | win_chance_Y |
|---|---|---|---|---|---|
| 18 | 1 | 0 | 0 | 1 | 1 |
| 20 | 0 | 1 | 0 | 0 | 1 |
| 23 | 0 | 1 | 0 | 1 | 1 |
| 19 | 1 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 1 | 1 | 0 |
| 22 | 0 | 1 | 0 | 1 | 1 |

(c)

[redacted]

**Feature construction (encoding nominal variables)**

[redacted]

## (b) Encoding Categorical (ordinal) variables

| | Science Marks | Maths Marks | Grade |
|---|---|---|---|
| ① | 78 | 75 | B |
| ② | 56 | 62 | C |
| ③ | 87 | 90 | A |
| ④ | 91 | 95 | A |
| ⑤ | 45 | 42 | D |
| ⑥ | 62 | 57 | B |

Table-1

| | Science Marks | Maths Marks | Grade |
|---|---|---|---|
| ① | 78 | 75 | 2 |
| ② | 56 | 62 | 3 |
| ③ | 87 | 90 | 1 |
| ④ | 91 | 95 | 1 |
| ⑤ | 45 | 42 | 4 |
| ⑥ | 62 | 57 | 2 |

Table-2

FC (c) Transforming Numeric (Cont.) features to Categorical features

| apartment_ area | apartment_ price |
|---|---|
| 4,720 | 23,60,000 |
| 2,430 | 12,15,000 |
| 4,368 | 21,84,000 |
| 3,969 | 19,84,500 |
| 6,142 | 30,71,000 |
| 7,912 | 39,56,000 |

(a)

| apartment_ area | apartment_ grade |
|---|---|
| 4,720 | Medium |
| 2,430 | Low |
| 4,368 | Medium |
| 3,969 | Low |
| 6,142 | High |
| 7,912 | High |

(b)

| apartment_ area | apartment_ grade |
|---|---|
| 4,720 | 2 |
| 2,430 | 1 |
| 4,368 | 2 |
| 3,969 | 1 |
| 6,142 | 3 |
| 7,912 | 3 |

(c)

**Feature construction (numeric to categorical)**

FC (d) Text - Specific Feature Construction :-

→ Text plays a major role in the flow of information, like for Social N/ws like FB, Twitter, Emails, SMS, Watsapp & hence become the most predominant medium of communication.

→ Text Mining is an imp. area of research - not only for technology practitioners but also for industry practitioners.

→ All ML models need numerical data as I/p, so text data in the data sets need to be transformed into numerical features.

→ Text data or Corpus which is the more popular keyword, is converted to a numerical representation by following a process k/as Vectorisation.

→ In this process, word occurrences in all documents belonging to the corpus are consolidated in the form of bag-of-words. There are 3 major steps:-

- a) Tokenize
  b) Count
  c) Normalize

→ In order to tokenize a corpus, the blank spaces and punctuations are used as delimiters to separate out the words, or tokens. Lastly tokens are weighted with reducing imp. when they occur in majority of documents.

→ A matrix is then formulated with each token representing a column & a specific document of the corpus representing each row.

→ Each cell contains the count of occurence of the token in a specific document. This matrix is k/as document-term matrix (also k/as term-document matrix).

Eg.

| S.No. | THIS | HOUSE | BUILD | FEELING | WELL | THEATRE | MOVIE | GOOD |
|---|---|---|---|---|---|---|---|---|
| ① | 2 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| ② | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| ③ | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 0 |
| ④ | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |

## ② Feature Extraction (FE):-

→ In feature Extraction, new features are created from a combination of original features.

→ Brian Ripley defined FE as follows:-

"FE is generally used to mean the construction of linear combinations of continous features which have good discriminatory power b/w classes."

→ Some of the commonly used operators for combining the original features include

① For Boolean features:- Conjunctions, Disconjunctions, Negation etc.

② For Nominal features:- Cartesian product, M of N, etc.

③ For Numerical features:- Min, Max, Addition, Subtraction, Multiplication, Dirision, Average, Equivalence, Inequality, etc.

Eg → Feature Set $F_i = F_1, F_2 \dots F_n$.

→ After feature extraction using a mapping function $f(F_1, F_2 \dots F_n)$ say we will have a set of features $F_i'(F_1', F_2' \dots F_m')$ such that $F_i' = f(F_i)$ & $m < n$

→ Eg $F_i' = K_1F_1 + K_2F_2$ :→ $F_1' = 0.3F_1 + 0.9F_4$ $F_2' = F_1 + 0.5F_2 + 0.6F_3$

| | $F_1$ | $F_2$ | $F_3$ | $F_4$ |
|---|---|---|---|---|
| ① | 34 | 34.5 | 23 | 233 |
| ② | 44 | 45.56 | 11 | 3.44 |
| ③ | 78 | 22.59 | 21 | 4.5 |
| ④ | 22 | 65.22 | 11 | 322.3 |
| ⑤ | 11 | 12.32 | 63 | 23.2 |

Table-1

| | $F_1'$ | $F_2'$ |
|---|---|---|
| ① | 41.25 | 185.80 |
| ② | 54.20 | 53.12 |
| ③ | 43.73 | 35.79 |
| ④ | 65.30 | 264.10 |
| ⑤ | 113.39 | 167.74 |

Table-2

⊛ The most popular feature extraction algorithm used in Machine Learning are:-

① Principal Component Analysis (PCA)
② Singular Value Decomposition (SVD)
③ Linear Discriminant Analysis (LDA)

## ⊛ Concept of Orthogonality:-

→ A vector space is a set of vectors
→ Vector spaces can be represented as a linear combination of a smaller set of vectors, called basis vectors. So, any vector 'v' in a vector space can be represented as

$$V = \sum_{i=1}^{n} a_i u_i$$

$a_i u_i$ ⟶ basis vectors
↳ represents 'n' scalars

→ Basis vectors are orthogonal to each other.
→ Orthogonality of vectors in n-dimensional vector space can be thought as an extension of vectors being perpendicular in a 2-dimensional vector space.
→ 2 orthogonal vectors are completely unrelated or independent of each other. So the transformation of a set of vectors to the corresponding set of basis vectors such that each vector in the original set can be expressed as a linear combination of basis vectors helps in decomposing the vectors to a no. of independent compone