In Semester Examination Artificial Intelligent and Machine Learning Question Paper for Government College of Engineering and Research, Avasari and its solution

Presentation · April 2022

DOI: 10.13140/RG.2.2.10692.30087

1 author:

Abhishek D. Patange
ABB

Mid-Semester Examination

Artificial Intelligent and Machine Learning

T. E. (Mechanical Engineering)

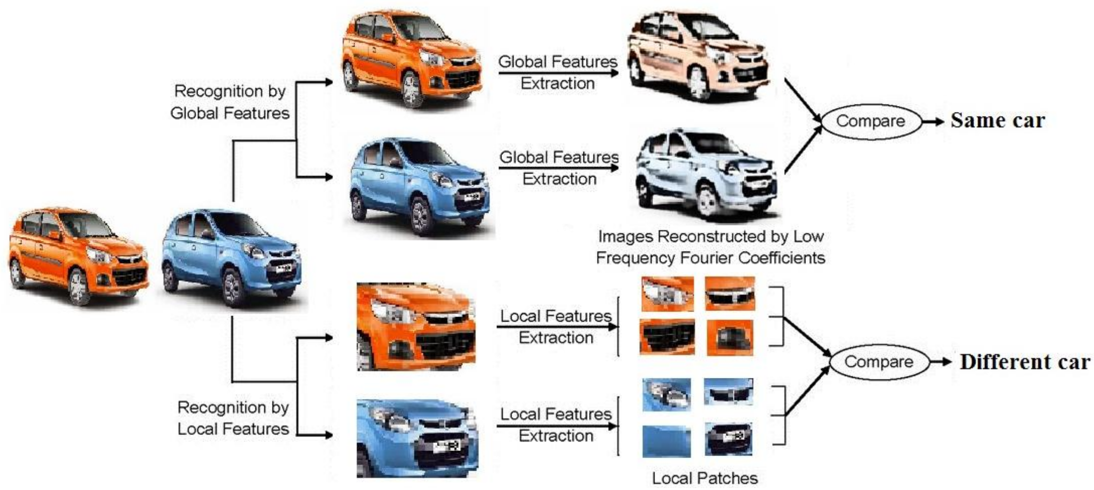
Government College of Engineering and Research Avasari (Kh) Pune

Time: 1.30 hours **Marks: 30**

1 Define and explain following terms with respect to example given below. 2

a. Global feature

b. Local feature



2 Represent over-fitting, under-fitting and optimum fitting in classification problem pictorially. State one real-life example. 4

3 Explain following terms used in PCA (Principal Components Analysis) algorithm. 4

a. Correlation b. Orthogonal

c. Eigenvectors d. Covariance Matrix

4 Explain supervised and unsupervised machine learning with one example. 4

OR

4 Explain following approaches to artificial intelligence. 4

a. Symbolic

b. Sub-symbolic

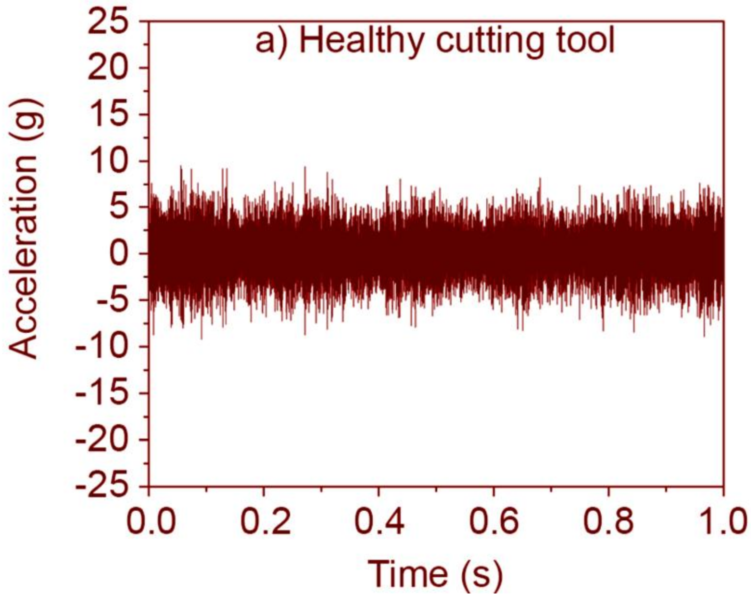
c. Statistical

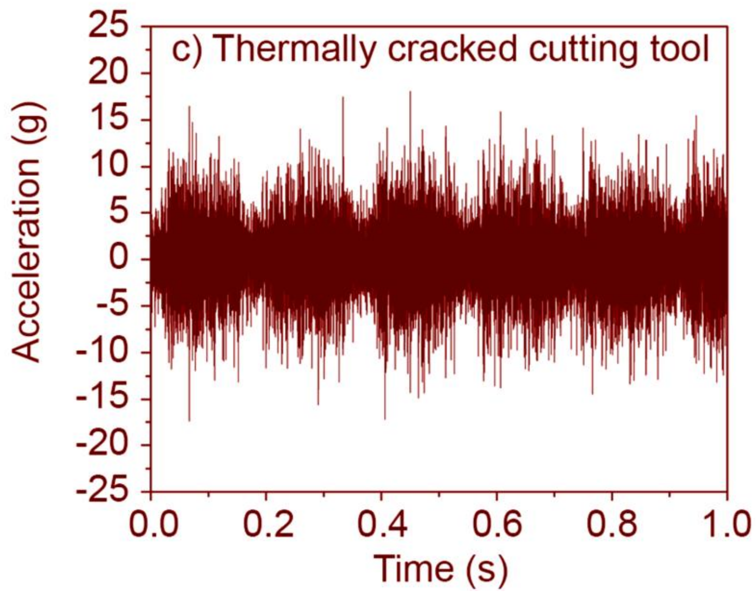
5 Differentiate between feature extraction and feature selection. 2

OR

5 Compare filter, wrapper and embedded methods. 2

6 Following graphs represent change in vibration signal (in terms of acceleration) with respect to time that depicts 2 conditions i.e. healthy and faulty cutting tool. 2

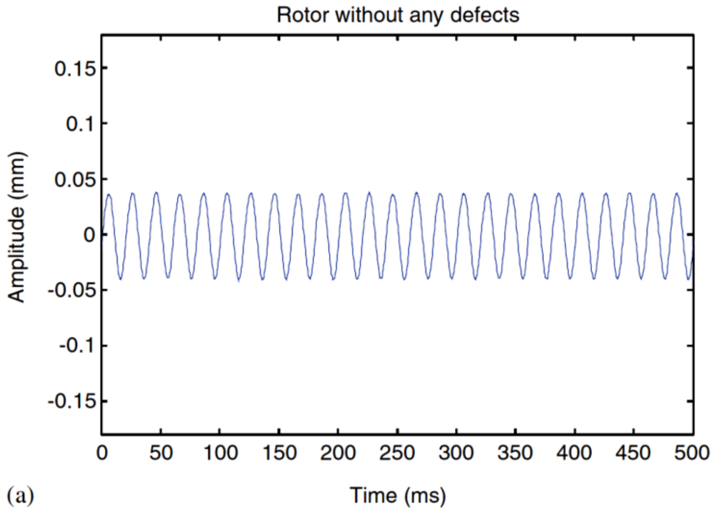




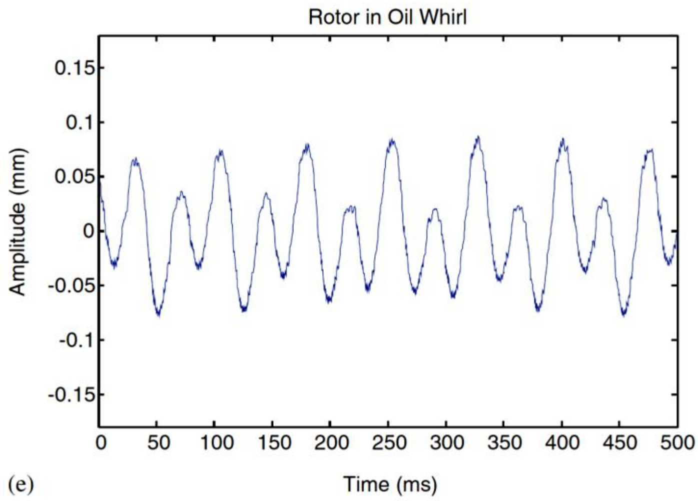
In order to develop machine learning based classification model, which statistical features will you extract from these vibration plots so as to depict difference between two.

OR

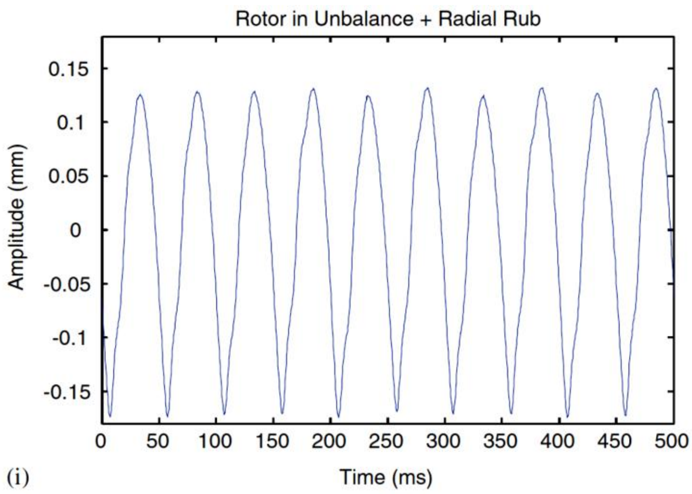
6 Following graphs represent change in displacement (in mm) with respect to time that depicts different conditions of rotor.



(a)



(e)



(i)

In order to develop machine learning based classification model, which statistical features will you extract from these vibration plots so as to depict difference between them.

5 Consider the following dataset which shows temperature of a thermal system at two different locations. 6

Temperature 1 (^{O}C)	2.5	0.5	2.2	1.9	3.1	2.3	2.0	1.0	1.5	1.1
Temperature 2 (^{O}C)	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

a. Standardize this data.
b. Find Eigen values and Eigen vectors.
c. Arrange Eigen values
d. Form feature vector
e. Transform original data and reconstruct it
f. Find principal components

6 Consider the training examples shown in following table below for a binary classification problem. 6

Instances	α_1	α_2	α_3	Target class (quality of steam generated by boiler)
1	T	T	1	Wet
2	T	T	6	Wet
3	T	F	5	Dry Saturated
4	F	F	4	Wet
5	F	T	7	Dry Saturated
6	F	T	3	Dry Saturated
7	F	F	8	Dry Saturated
8	T	F	7	Wet
9	F	T	5	Dry Saturated

a. What are the information gains of α_1 and α_2 relative to these training examples?
b. For α_3 which is a continuous attribute, compute the information gain for every possible split.
c. What is the best split (among α_1, α_2 and α_3) according to the information gain?

****

Mid-Semester Examination

Artificial Intelligent and Machine Learning

T. E. (Mechanical Engineering)

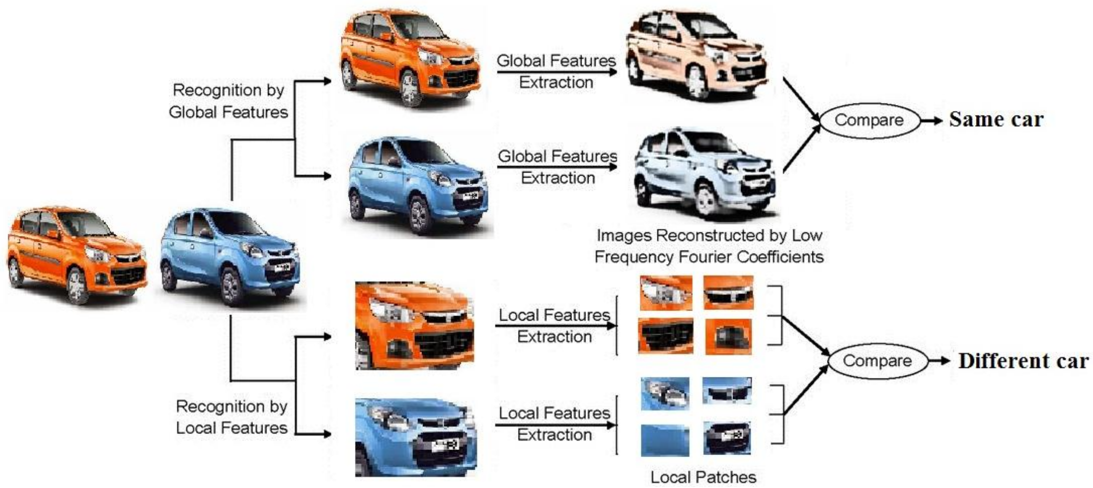
Government College of Engineering and Research Avasari (Kh) Pune

Time: 1.30 hours **Marks: 30**

1 Define and explain following terms with respect to example given below. 2

c. Global feature

d. Local feature



First of all definition of Global and Local feature is expected:

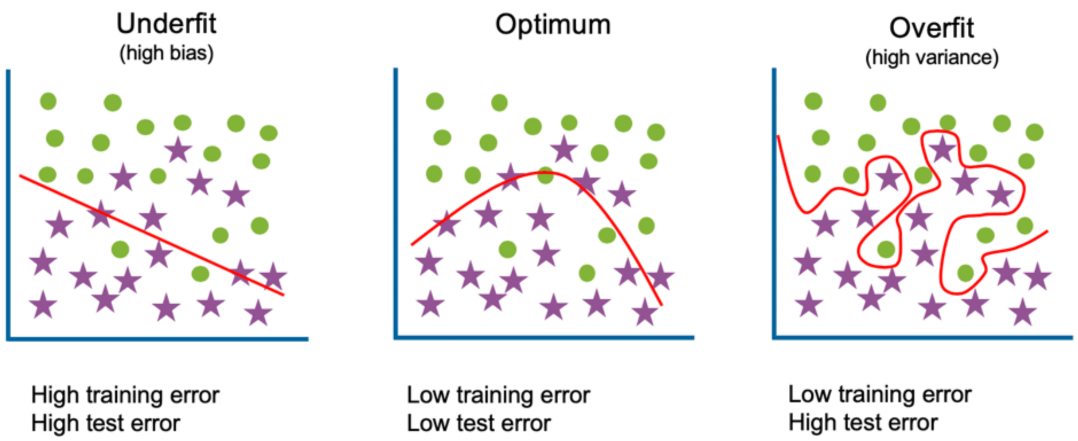
- Global features describe the visual content of the entire image by a single vector. They represent the texture, color, shape information which are the most popular for image representation.
- Local features aim to detect the interest points (IPs) in an image and describe them by a set of vectors.
- Simply speaking, global features describe the entire image, whereas local features describe the image patches (small group of pixels).

Explanation with respect to example:

- Here in the comparison of 2 images of Maruti-Suzuki Alto K10 and Alto 800, global features describe the texture, shape information of the entire image of Maruti-Suzuki Alto K10 and Alto 800 by a single vector.
- So overall both bikes seems to be same. For this example, using global features representation may create confusion.
- On the other hand local features i.e. image patches (front end of Maruti-Suzuki Alto K10 & Alto 800) in terms of small group of pixels tries to detect interest points (IPs).

This would help in identifying difference between two cars. Thus local features seem to be useful to develop a classification model.

2 Represent over-fitting, under-fitting and optimum fitting in classification problem pictorially. State one real-life example. 4

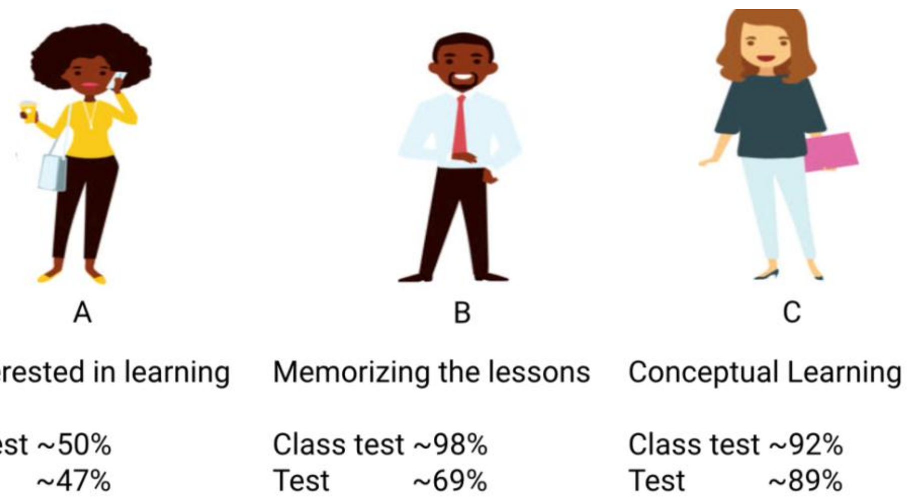


Example: predict who will be an "A" student in college.

- Let's say "Peter Evans Hope" is an "A" student in your dataset.
- If your model says anyone named "Peter Evans Hope" is an "A" student, the model will correctly predict this specific student but it is over-fitted.
- Because in the general population, there probably isn't another "Peter Evans Hope" (and if there is, probably not an "A" student).
- If your model says anyone who graduated from [insert top high school] will be an "A" student in college, this is under-fitting (too general).

Within the graduates from that high school, there will be a range of college GPAs - what else can explain who gets the "A"?

OR



3 Explain following terms used in PCA (Principal Components Analysis) algorithm. 4

b. Correlation b. Orthogonal

d. Eigenvectors d. Covariance Matrix

- **Correlation:** It signifies that how strongly two variables are related to each other. Such as if one changes, the other variable also gets changed. The correlation value ranges from -1 to +1. Here, -1 occurs if variables are inversely proportional to each other, and +1 indicates that variables are directly proportional to each other.
- **Orthogonal:** It defines that variables are not correlated to each other, and hence the correlation between the pair of variables is zero.
- **Eigenvectors:** If there is a square matrix M, and a non-zero vector v is given. Then v will be eigenvector if Av is the scalar multiple of v.
- **Covariance Matrix:** A matrix containing the covariance between the pair of variables is called the Covariance Matrix.

4 Explain supervised and unsupervised machine learning with one example. 4

Supervised Machine Learning:

Supervised learning is a machine learning method in which models are trained using labeled data. In supervised learning, models need to find the mapping function to map the input variable (X) with the output variable (Y).

$y = f(x)$

Supervised learning needs supervision to train the model, which is similar to as a student learns things in the presence of a teacher. Supervised learning can be used for two types of problems: **Classification** and **Regression**.

Example: Suppose we have an image of different types of fruits. The task of our supervised learning model is to identify the fruits and classify them accordingly. So to identify the image in supervised learning, we will give the input data as well as output for that, which means we will train the model by the shape, size, color, and taste of each fruit. Once the training is completed, we will test the model by giving the new set of fruit. The model will identify the fruit and predict the output using a suitable algorithm.

Unsupervised Machine Learning:

Unsupervised learning is another machine learning method in which patterns inferred from the unlabelled input data. The goal of unsupervised learning is to find the structure and patterns from the input data. Unsupervised learning does not need any supervision. Instead, it finds patterns from the data by its own.

Unsupervised learning can be used for two types of problems: **Clustering** and **Association**.

Example: To understand the unsupervised learning, we will use the example given above. So unlike supervised learning, here we will not provide any supervision to the model. We will just provide the input dataset to the model and allow the model to find the patterns from the data. With the help of a suitable algorithm, the model will train itself and divide the fruits into different groups according to the most similar features between them.

OR

4 Explain following approaches to artificial intelligence. 4

d. Symbolic
e. Sub-symbolic
f. Statistical

Symbolic AI refers to the fact that all steps are based on symbolic human readable representations of the problem that use logic and search to solve problem. Key advantage of Symbolic AI is that the reasoning process can be easily understood – a Symbolic AI program can easily explain why a certain conclusion is reached and what the reasoning steps had been. A key disadvantage of Symbolic AI is that for learning process – the rules and knowledge has to be hand coded which is a hard problem. So far, symbolic AI has been confined to the academic world and university labs with little research coming from industry giants.

Non-symbolic AI performs calculations according to some principles that have demonstrated to be able to solve problems. Without exactly understanding how to arrive at the solution. Examples of Non-symbolic AI include genetic algorithms, neural networks and deep learning. The origins of non-symbolic AI come from the attempt to mimic a human brain and its complex network of interconnected neurons. Non-symbolic AI is also known as "Connectionist AI" and the current applications are based on this approach – from Google's automatic transition system (that looks for patterns), IBM's Watson, Facebook's face recognition algorithm to self-driving car technology. A key disadvantage of Non-symbolic AI is that it is difficult to understand how the system came to a conclusion. This is particularly important when applied to critical applications such as self-driving cars, medical diagnosis among others. Non-symbolic systems such as DL-powered applications cannot take high-risk decisions.

5 Differentiate between feature extraction and feature selection. 2

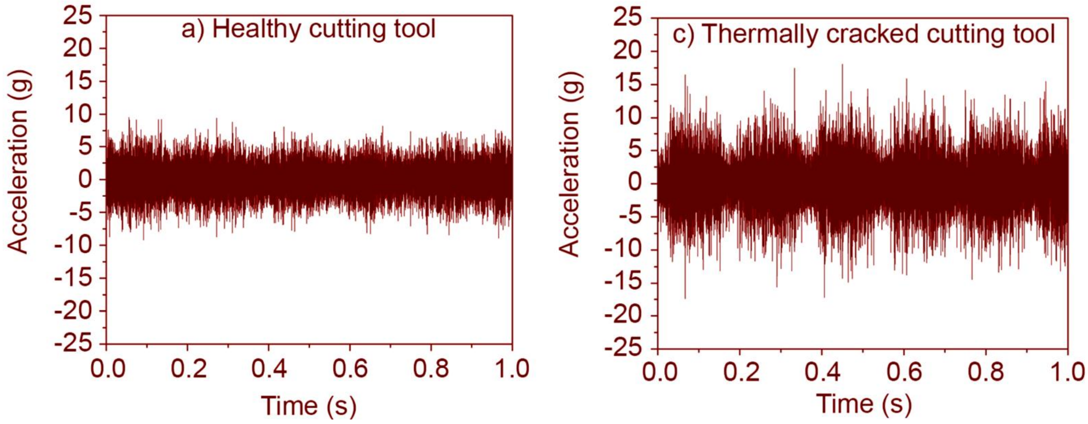
The **key difference** between feature selection and feature extraction techniques used for dimensionality reduction is that while the **original features are maintained** in the case of feature selection algorithms, the feature extraction algorithms **transform the data onto a new feature space**. Feature selection techniques can be used if the requirement is to **maintain the original features,** unlike the feature extraction techniques which derive useful information from data to construct a new feature subspace. Feature selection techniques are used when model explainability is a key requirement. Feature extraction techniques can be used to improve the predictive performance of the models, especially, in the case of **algorithms that don't support regularization**. Unlike feature selection, feature extraction usually needs to transform the original data to features with strong pattern recognition ability, where the original data can be regarded as features with weak recognition ability.

OR

5 Compare filter, wrapper and embedded methods. 2

Filter methods	Wrapper methods	Embedded methods
Generic set of methods which do not incorporate a **specific machine learning algorithm.**	Evaluates on a **specific machine learning algorithm** to find optimal features.	Embeds (fix) features during **model building process.** Feature selection is done by observing each iteration of model training phase.
Much **faster** compared to Wrapper methods in terms of time complexity	**High computation time** for a dataset with many features	Sits **between Filter methods and Wrapper methods** in terms of time complexity
Less prone to **over-fitting**	High chances of **over-fitting** because it involves training of machine learning models with different combination of features	Generally used to reduce **over-fitting** by **penalizing** the coefficients of a model being too large.
Examples – **Correlation, Chi-Square test, ANOVA, Information gain** etc.	Examples - **Forward Selection, Backward elimination, Stepwise selection** etc.	Examples - **LASSO, Elastic Net, Ridge Regression** etc.

6 Following graphs represent change in vibration signal (in terms of acceleration) with respect to time that depicts 2 conditions i.e. healthy and faulty cutting tool. 2

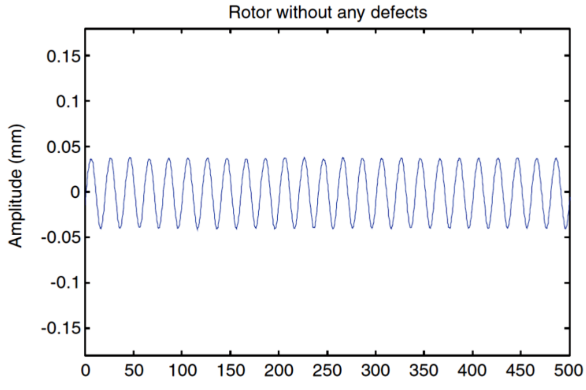


In order to develop machine learning based classification model, which statistical features will you extract from these vibration plots so as to depict difference between two.

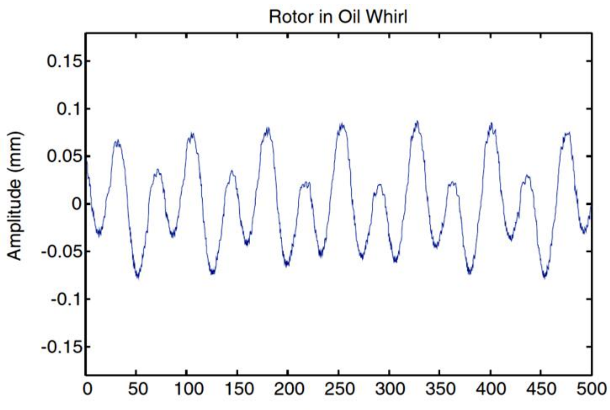
Features such as mean, mode, median, kurtosis, skewness, standard deviation, variance etc. can be extracted from both the signals in order to examine variation between the classes. Let's take one simple feature as maximum value of acceleration. If we extract maximum value of acceleration for both conditions of tool i.e. healthy and thermally cracked, it is evident that the maximum value of acceleration is less for healthy tool condition however it appears to be increased for the case of thermally cracked tool. Thus a maximum value feature is useful to develop a classification model.

OR

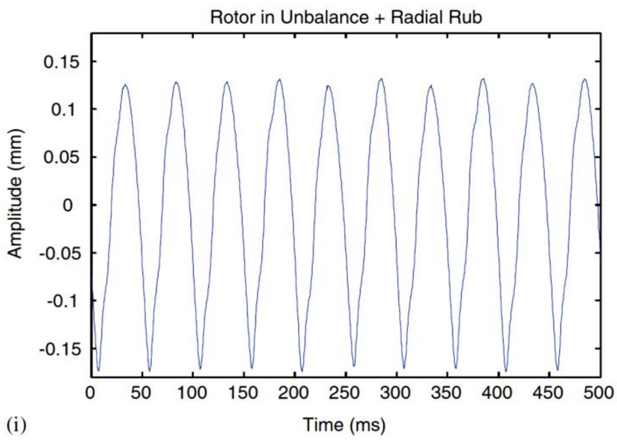
6 Following graphs represent change in displacement (in mm) with respect to time that depicts different conditions of rotor.



(a)



(e)



(i)

In order to develop machine learning based classification model, which statistical features will you extract from these vibration plots so as to depict difference between them.

Features such as mean, mode, median, kurtosis, skewness, standard deviation, variance etc. can be extracted from both the signals in order to examine variation between the classes. Let's take one simple feature as maximum value of displacement. If we extract maximum value of displacement for all three cases, it is evident that the maximum value of displacement is less for rotor without any defect however it appears to be increased for the other two cases i.e. rotor in oil whirl and rotor in unbalance + radial rub. Thus a maximum value feature is useful to develop a classification model.

5 Consider the following dataset which shows temperature of a thermal system at two different locations. 6

Temperature 1 (^{O}C)	2.5	0.5	2.2	1.9	3.1	2.3	2.0	1.0	1.5	1.1
Temperature 2 (^{O}C)	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9

g. Standardize this data.
h. Find Eigen values and Eigen vectors.
i. Arrange Eigen values
j. Form feature vector
k. Transform original data and reconstruct it
l. Find principal components

Step 1: Standardize the Dataset

Mean for x_1= 1.81 = x_{1mean}

Mean for x_2= 1.91 = x_{2mean}

We will change the dataset.

x1	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
x2	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01

Step 2: Find the Eigenvalues and eigenvectors

Correlation Matrix c = $C = \left(\frac{X \cdot X^T}{N-1}\right)$

where, X is the Dataset Matrix (In this numerical, it is a 10 X 2 matrix)

X^T is the transpose of the X (In this numerical, it is a 2 X 10 matrix) and N is the number of elements = 10

So, $C = \left(\frac{X \cdot X^T}{10-1}\right) = \left(\frac{X \cdot X^T}{9}\right)$

{So in order to calculate the Correlation Matrix, we have to do the multiplication of the Dataset Matrix with its transpose}

$$C = \begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix}$$

Using the equation, **| C – λI | = 0– equation (i)** where { \lambda is the eigenvalue and I is the Identity Matrix }

So solving equation (i)

$$\begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{vmatrix} 0.616556 - \lambda & 0.615444 \\ 0.615444 & 0.716556 - \lambda \end{vmatrix} = 0$$

Taking the determinant of the left side, we get

$$0.44180 - 0.616556\lambda - 0.716556\lambda + \lambda^2 - 0.37877 = 0$$

$$\lambda^2 - 1.33311\lambda + 0.06303 = 0$$

We get two values for λ, that are **(λ_1) = 1.28403 and (λ_2) = 0.0490834**. Now we have to find the eigenvectors for the eigenvalues λ_1 **and** λ_2

To find the eigenvectors from the eigenvalues, we will use the following approach:

First, we will find the eigenvectors for the eigenvalue 1.28403 by using the equation $C \cdot X = \lambda \cdot X$

$$\begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 1.28403 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} 0.616556x + 0.615444y \\ 0.615444x + 0.716556y \end{bmatrix} = \begin{bmatrix} 1.28403x \\ 1.28403y \end{bmatrix}$$

Solving the matrices, we get

0.616556x + 0.615444y = 1.28403x ; **x = 0.922049 y**

(x and y belongs to the matrix X) so if we put y = 1, x comes out to be 0.922049. So now the updated X matrix will look like:

$$X = \begin{bmatrix} 0.922049 \\ 1 \end{bmatrix}$$

IMP: Till now we haven't reached to the eigenvectors, we have to a bit of modifications in the X matrix. They are as follows:

A. Find the square root of the sum of the squares of the element in X matrix i.e.

$$\sqrt{0.922049^2 + 1^2} = \sqrt{0.850174 + 1} = \sqrt{1.850174} = 1.3602$$

B. Now divide the elements of the X matrix by the number 1.3602 (just found that)

$$\begin{bmatrix} \frac{0.922049}{1.3602} \\ \frac{1}{1.3602} \end{bmatrix} = \begin{bmatrix} 0.67787 \\ 0.73518 \end{bmatrix}$$

So now we found the eigenvectors for the eigenvector λ_1, they are 0.67787 and 0.73518

Secondly, we will find the eigenvectors for the eigenvalue 0.0490834 by using the equation {Same approach as of previous step)

$$\begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 0.0490834 \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} 0.616556x + 0.615444y \\ 0.615444x + 0.716556y \end{bmatrix} = \begin{bmatrix} 0.0490834x \\ 0.0490834y \end{bmatrix}$$

Solving the matrices, we get

0.616556x + 0.615444y = 0.0490834x; **y = -0.922053**

(x and y belongs to the matrix X) so if we put x = 1, y comes out to be -0.922053 So now the updated X matrix will look like:

$$X = \begin{bmatrix} 1 \\ -0.922053 \end{bmatrix}$$

IMP: Till now we haven't reached to the eigenvectors, we have to a bit of modifications in the X matrix. They are as follows:

A. Find the square root of the sum of the squares of the elements in X matrix i.e.

$$\sqrt{1^2 + (-0.922053)^2} = \sqrt{1 + 0.85018} = \sqrt{1.85018} = 1.3602$$

B. Now divide the elements of the X matrix by the number 1.3602 (just found that)

$$\begin{bmatrix} \frac{1}{1.3602} \\ \frac{-0.922053}{1.3602} \end{bmatrix} = \begin{bmatrix} 0.735179 \\ 0.677873 \end{bmatrix}$$

So now we found the eigenvectors for the eigenvector \lambda_2, they are 0.735176 and 0.677873

Sum of eigenvalues (λ_1) and (λ_2) = 1.28403 + 0.0490834 = 1.33 = Total Variance {Majority of variance comes from λ_1}

Step 3: Arrange Eigenvalues

The eigenvector with the highest eigenvalue is the Principal Component of the dataset. So in this case, eigenvectors of lambda1 are the principal components.

{Basically in order to complete the numerical we have to only solve till this step, but if we have to prove why we have chosen that particular eigenvector we have to follow the steps from 4 to 6}

Step 4: Form Feature Vector

$\begin{bmatrix} 0.677873 & 0.735179 \\ 0.735179 & -0.677879 \end{bmatrix}$ This is the FEATURE VECTOR for Numerical

Where first column are the eigenvectors of λ_1 & second column are the eigenvectors of λ_2

Step 5: Transform Original Dataset

Use the equation Z = X V

$$\begin{bmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{bmatrix} \cdot \begin{bmatrix} 0.677873 & 0.735179 \\ 0.735179 & -0.677879 \end{bmatrix} = \begin{bmatrix} 0.8297008 & 0.17511574 \\ -1.77758022 & -0.14285816 \\ 0.99219768 & -0.38437446 \\ 0.27421048 & -0.13041706 \\ 1.67580128 & 0.20949934 \\ 0.91294918 & -0.17528196 \\ -1.14457212 & -0.04641786 \\ -0.43804612 & -0.01776486 \\ -1.22382.62 & 0.16267464 \end{bmatrix} = Z$$

Step 6: Reconstructing Data

Use the equation X = $Z * V^T$ ($V^T$is Transpose of V), X = Row Zero Mean Data

$$\begin{bmatrix} 0.8297008 & 0.17511574 \\ -1.77758022 & -0.14285816 \\ 0.99219768 & -0.38437446 \\ 0.27421048 & -0.13041706 \\ 1.67580128 & 0.20949934 \\ 0.91294918 & -0.17528196 \\ -1.14457212 & -0.04641786 \\ -0.43804612 & -0.01776486 \\ -1.22382.62 & 0.16267464 \end{bmatrix} \cdot \begin{bmatrix} 0.677873 & 0.735179 \\ 0.735176 & -0.677879 \end{bmatrix} = \begin{bmatrix} 0.6899999766573 & 0.4899999834233 \\ -1.3099999556827 & -1.2099999590657 \\ 0.389999968063 & 0.9899999665083 \\ 0.0899999969553 & 0.2899999901893 \\ 0.61212695653593 & 0.35482096313253 \\ 0.4899999834233 & 0.7899999732743 \\ 0.189999935723 & -0.309999995127 \\ -0.8099999725977 & -0.8099999725977 \\ -0.3099999895127 & -0.3099999895127 \\ -0.7099999759807 & -1.0099999658317 \end{bmatrix}$$

So in order to reconstruct the original data, we follow:

Row Original DataSet = Row Zero Mean Data + Original Mean

$$\begin{bmatrix} 0.6899999766573 & 0.4899999834233 \\ -1.3099999556827 & -1.2099999590657 \\ 0.389999968063 & 0.9899999665083 \\ 0.0899999969553 & 0.2899999901893 \\ 0.61212695653593 & 0.35482096313253 \\ 0.4899999834233 & 0.7899999732743 \\ 0.189999935723 & -0.309999995127 \\ -0.8099999725977 & -0.8099999725977 \\ -0.3099999895127 & -0.3099999895127 \\ -0.7099999759807 & -1.0099999658317 \end{bmatrix} + \begin{bmatrix} 1.81 & 1.91 \end{bmatrix} = \begin{bmatrix} 2.49 & 2.39 \\ 0.5 & 0.7 \\ 2.19 & 2.89 \\ 1.89 & 2.19 \\ 3.08 & 2.99 \\ 2.30 & 2.7 \\ 2.01 & 1.59 \\ 1.01 & 1.11 \\ 1.5 & 1.6 \\ 1.1 & 0.9 \end{bmatrix}$$

So for the eigenvectors of first eigenvalue, data can be reconstructed similar to the original dataset. Thus we can say that the Principal Component of the dataset is $\lambda_1$is 1.28403 followed by λ_2**that is 0.0490834**

6 Consider the training examples shown in following table below for a binary classification problem. 6

Instances	α_1	α_2	α_3	Target class (quality of steam generated by boiler)
1	T	T	1	Wet
2	T	T	6	Wet
3	T	F	5	Dry Saturated
4	F	F	4	Wet
5	F	T	7	Dry Saturated
6	F	T	3	Dry Saturated
7	F	F	8	Dry Saturated
8	T	F	7	Wet
9	F	T	5	Dry Saturated

d. What are the information gains of α_1 and α_2 relative to these training examples?

e. For α_3 which is a continuous attribute, compute the information gain for every possible split.

f. What is the best split (among α_1, α_2 and α_3) according to the information gain?

(a) What is the entropy of this collection of training examples with respect to the positive class?

Answer:

There are four positive examples and five negative examples. Thus, $P(+) = 4/9$ and $P(-) = 5/9$. The entropy of the training examples is $-4/9 \log_2(4/9) - 5/9 \log_2(5/9) = 0.9911$.

(b) What are the information gains of a_1 and a_2 relative to these training examples?

Answer:

For attribute a_1, the corresponding counts and probabilities are:

a_1	+	-
T	3	1
F	1	4

The entropy for a_1 is

$$\frac{4}{9}\left[-(3/4)\log_2(3/4)-(1/4)\log_2(1/4)\right] + \frac{5}{9}\left[-(1/5)\log_2(1/5)-(4/5)\log_2(4/5)\right]=0.7616.$$

Therefore, the information gain for a_1 is $0.9911-0.7616=0.2294$.

For attribute a_2, the corresponding counts and probabilities are:

a_2	+	-
T	2	3
F	2	2

The entropy for a_2 is

$$\frac{5}{9}\left[-(2/5)\log_2(2/5)-(3/5)\log_2(3/5)\right] + \frac{4}{9}\left[-(2/4)\log_2(2/4)-(2/4)\log_2(2/4)\right]=0.9839.$$

Therefore, the information gain for a_2 is $0.9911-0.9839=0.0072$.

(d) What is the best split (among a_1, a_2, and a_3) according to the information gain?

Answer:

According to information gain, a_1 produces the best split.

****