

MACHINE LEARNING IN SIGNAL PROCESSING

MIN MAX PROCEDURE

$$= \left(\frac{V - \text{min}}{\text{max} - \text{min}} \right) \times (\text{new max} - \text{new min}) + \text{new min}$$

Z-SCORE NORMALIZATION

$$Z = \frac{V - \mu}{\sigma}, \quad \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$$

eg] $V = \{88, 90, 92, 94\}$

Quartiles & Inter Quartile Range

It is sometimes convenient to subdivide dataset using co-ordinates. Percentiles are about data that are less than the coordinate by some % of the total value i.e., k th percentile is the property that the $k\%$ data lies at or below x_i . For example, median is 3rd percentile and can be denoted as $Q_{0.5}$. The 25th percentile is called the first quartile (Q_1) and the 75th percentile is called third quartile (Q_3). Another measure that is useful to measure dispersion is Inter quartile range (IQR).

$$\text{IQR} = Q_3 - Q_1$$

Outliers are normally the values falling apart at least by the amount $1.5 \times \text{IQR}$ above the third quartile or below first quartile.

Inter quartile is defined by $Q_{0.75} - Q_{0.25}$.

eg: For patient's age $\{12, 14, 19, 22, 24, 26, 28, 31, 34\}$

Soln: 24 is median

The first quartile is median of the scores below 24.

$$\text{median of } \{12, 14, 19, 22\} \Rightarrow \frac{14+19}{2} = 16.5$$

$$\downarrow \\ Q_{0.25}$$

The 3rd quartile is $\{26, 28, 31, 34\}$

$$\text{median} \Rightarrow \frac{28+31}{2} = 29.5$$

$$Q_{0.75}$$

$$T_{0.25} = 29.5 - 16.5 = 13$$



CLINIQUE

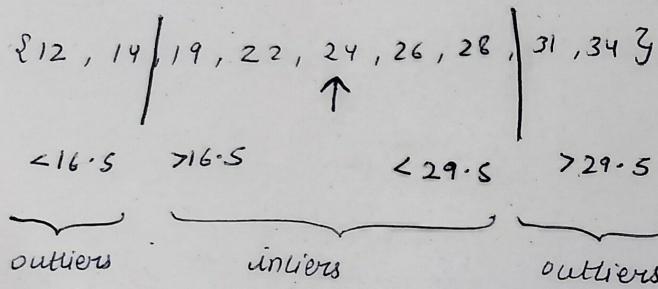
MAYBELLINE
INSTANT

The half of Inter Quartile Range is called semi Inter Quartile Range.

$$SIQR = \frac{1}{2} IQR$$

$$\therefore SIQR = \frac{1}{2} \times 13 = 6.5$$

dividing Quartile with $Q_{0.25}$ and $Q_{0.75}$ as dividing no.,



$$\text{eg: } \{20, 10, 30\} \rightarrow \mu = 20$$

$$\{60, 0, 0\} \rightarrow \mu = 20$$

$$\{20, 40, 0\} \rightarrow \mu = 20$$

$$\{10, 0, 50\} \rightarrow \mu = 20$$

$$SD = \sqrt{\frac{\sum_{i=1}^N (u_i - \bar{x})^2}{N-1}} \Rightarrow \sigma_1 = \sqrt{\frac{(20-20)^2 + (10-20)^2 + (30-20)^2}{3-1}} = \sqrt{\frac{200}{2}} = 10$$

$$\sigma_2 = \sqrt{\frac{(60-20)^2 + (10-20)^2 + (0-20)^2}{3-1}} = 20\sqrt{3}$$

more the variance, easier the classification.
standard deviation

Generic Algorithm for feature selection

GA is heuristic search algorithm based on the Charles Darwin's Theory of survival of the fittest. In this algorithm, an initial population created which consists of many ~~characteristic~~ chromosomes, each of which represents a possible solution to a problem.

There is a fitness function which assigns a fitness value to each chromosome. This fitness value represents the goodness of the solution. Initially the population may or may not contain good solution. In GA, each iteration generates a new chromosome by operations like crossover and mutation.

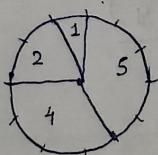
The crossover carried out by choosing two chromosomes and creating a new chromosome out of them. There can be many cross-overs. For eg: one point crossover, two-point crossover, multi-point crossover.

Mutation on other hand, is the process of creating mutants.

For eg: selecting a binary chromosome and inverting a particular cell of the chromosome can be one way of creating mutations mutants. The mutation is generally used to break the local maxima. The selection is carried out by techniques like Roulette wheel selection. In this method, the probability of selecting chromosomes with a higher fitness value is more.

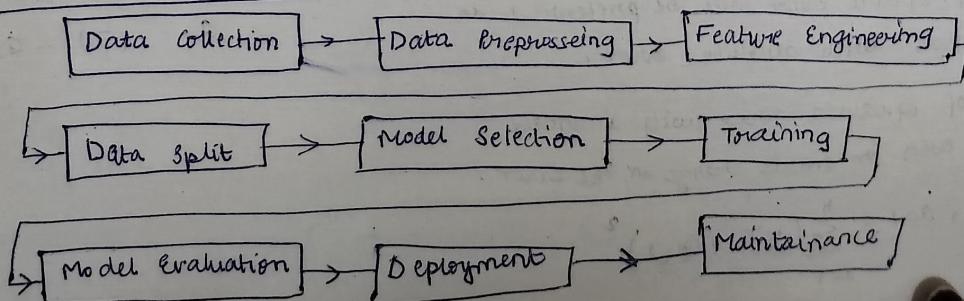
If four chromosomes have fitness values [4, 2, 1, 5] then pie chart with sectors having values can be computed.

$$\frac{4}{4+2+1+5} \times 360^\circ = 120^\circ, \quad \frac{2}{12} \times 360^\circ = 60^\circ, \quad \frac{1}{12} \times 360^\circ = 30^\circ, \quad \frac{5}{12} \times 360^\circ = 150^\circ$$



5 is the fittest value %. that chromosome will be selected.

Machine Learning Pipeline



Linear Regression

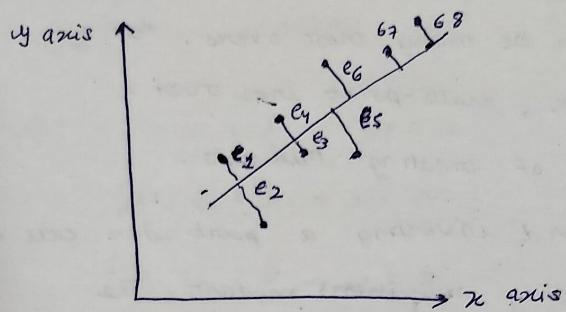
The linear regression model can be created by fitting a line among scattered data points.

$$y = a_0 + a_1 x + e$$

Here a_0 is the 'intercept' which represents bias and a_1 represents the slope of line. a_0, a_1 are called regression coefficients. e is called the error in predictions.

The assumptions of linear regression are

- The observations (y_i) are random and are mutually independent.
- The error is also mutually independent and has some distribution.



$$y_1 = (a_0 + a_1 x_1) + e_1$$

$$y_2 = (a_0 + a_1 x_2) + e_2$$

⋮

$$y_n = (a_0 + a_1 x_n) + e_n$$

$$e = y_i - (a_0 + a_1 x_i)$$

Error

$$E = \sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - (a_0 + a_1 x_i))$$

minimization of absolute value of error

$$E = \sum_{i=1}^n |e_i| = \sum_{i=1}^n |y_i - (a_0 + a_1 x_i)|$$

minimization of sum of square errors

$$E = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (y_i - (a_0 + a_1 x_i))^2$$

Sum of square error will be preferred in comparison with absolute error.

Sum of squared error will increase large even for small change in the error.

$$J(a_1, a_0) = \sum_{i=1}^n (y_i - f(x_i))^2$$

$$= \sum_{i=1}^n [y_i - (a_0 + a_1 x_i)]^2$$

$J(a_0, a_1)$ is the function of parameters a_0 and a_1 .

This needs to be minimized by differentiating and making equals to zero. This yields the coefficient values of a_0 and a_1 .

$$a_1 = \frac{\bar{xy} - (\bar{x})(\bar{y})}{(\bar{x}^2) - (\bar{x})^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

eg1: Let us consider 5 weeks sales data is given in below table. Apply linear regression to find 7th and 9th week sales data.

| x_i | y_i | xy | x_i^2 |
|-------|-------|------|---------|
| 1 | 1.2 | 1.2 | 1 |
| 2 | 1.8 | 3.6 | 4 |
| 3 | 2.6 | 7.8 | 9 |
| 4 | 3.2 | 12.8 | 16 |
| 5 | 3.8 | 19 | 25 |

$$\bar{x} = 3 \quad \bar{y} = 2.52 \quad \bar{xy} = 8.85 \quad \bar{x}^2 = 1$$

$$a_1 = \frac{8.85 - (3)(2.52)}{11 - 9} = 0.66$$

$$a_0 = 2.52 - (0.66)(3) = 0.54$$

$$\begin{aligned} \hat{y}_7 &= a_0 + a_1 x_7 \\ &= 0.54 + 0.66 \times 7 = 5.16 \end{aligned}$$

$$\begin{aligned} \hat{y}_9 &= a_0 + a_1 x_9 = 0.54 + 0.66 \times 9 \\ &= 6.48 \end{aligned}$$

Q] Find the linear regression of data given below, predict stock value after 90 days

Relative MSE

$$\textcircled{5} \quad \text{MSE} = \frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2}$$

↑ predicted value
↓ mean of y_i

Root mean square error

$$\textcircled{4} \quad \text{RMSE} = \sqrt{\text{MSE}}$$

$$= \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}$$

\textcircled{3} Mean squared error (MSE)

$$= \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2$$

\textcircled{6} coefficient of variation

$$C = \frac{\text{RMSE}}{\bar{y}}$$

Z - Test

Z test assumes normal distribution of data whose population variation is known. The sample size assumed to be large. The focus is to test the population mean.

The Z-statistic is

$$Z = \frac{x - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

Here, x → input data

n → no. of data elements

μ → mean of x

σ → standard deviation of x

t-test and paired t-test

t-test is a hypothesis test and checks if the difference between two sample mean is real or by chance. Here data is continuous and randomly selected. There will be only small number of samples and variance between groups is real.

The t-test statistic follows t-distribution under null hypothesis and is used when the number of samples less than 30 or small in number.

The distribution follows t-distribution rather than Gaussian distribution. It indicates two groups are different or not.

One sample test

In this test, the mean of the two group is checked against the set average that can be either theoretical value or population mean, so the procedure is

- 1) Select a group
- 2) Compute average
- 3) Compute it with theoretical value and compute t-statistic.

$$t = \frac{m - u}{\frac{s}{\sqrt{n}}}$$

t → is t statistic

m → mean of group

u → theoretical value or population mean

s → standard deviation

n → the group size or sample size

Singular values and Singular Vectors in SVD

The best matrices (real symmetric matrices S) have real eigen values and orthogonal eigen vectors. But for other matrices the eigen values are complex or the eigen values are not orthogonal.

If A is not square then $Ax = \lambda x$ is impossible and eigen vectors fail.
(left side in R^m and right side in R^n)

We need an idea to succeed for every matrix.

The singular value decomposition fills this gap in a perfect way. In our applications A is often a matrix of data. the rows could tell us the age and height of 1000 children, then A is 2 by 1000. definitely rectangular, unless height is exactly proportional to age, the rank is $r=2$ and that matrix A has two positive singular values of σ_1 and σ_2 .

Singular values and singular vectors in SVD

The key point is that we need two sets of singular vectors. The u 's and v 's from a real m by n matrix, the ' n ' right singular vectors v_1, v_2, \dots, v_n are orthogonal in \mathbb{R}^m .

The m left singular vectors u_1, u_2, \dots, u_m are perpendicular to each other in \mathbb{R}^n . The connection between n v 's and m u 's is not $Au = \sigma v$. That is for eigen vectors for singular vectors: each Au equals σv :

$$Av_1 = \sigma_1 u_1, \dots, Av_n = \sigma_n u_n$$

$$Av_{n+1} = 0, \dots, Av_m = 0 \quad \text{--- (1)}$$

I have separated the first r v 's and u 's from the rest. The number r is the rank of A . The number of independent columns (and rows). Then r is the dimension of the column space and the row space.

We will have r positive singular values in descending order.

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \dots \sigma_r > 0$$

The last $n-r$ v 's are null space of A and last $m-r$ u 's are null space of A^T .

Our first step is to write (1) in matrix form. All of the right singular vectors v_1 to v_n go in the columns of V .

The left singular vectors u_1 to u_m go in columns of U .

These are square orthogonal matrices $V^T = V^{-1}$ and $U^T = U^{-1}$ because their columns are orthogonal unit vectors.

Then eqn (1) becomes the full ~~SVD~~ SVP with square matrices V and U .

$$AV = U\Sigma$$

$$A \begin{bmatrix} v_1 v_2 \dots v_n \end{bmatrix} = \begin{bmatrix} u_1 u_2 \dots u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \\ & & & \sigma_r \\ 0 & & & 0 \end{bmatrix}$$

You see $Av_k = \sigma_k u_k$ in the first m columns above.

That is the important part of SVP. It shows the basis of v 's for the row space of A and then u 's for column space. After positive numbers $\sigma_1, \sigma_2, \dots, \sigma_r$ or the main diagonal of Σ the rest of that matrix is all zero from the null spaces of A and A^T .

The eigen vectors give $AX = XA$.

But $AV = U\Sigma$ needs two sets of singular vectors.

Example ①

$$AV = U\Sigma \quad \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3\sqrt{n} \\ \sqrt{n} \end{bmatrix}$$

The matrix A is not symmetric, so V is different from U . The rank is 2, so there are two singular values, $\sigma_1 = 3\sqrt{n}$ and $\sigma_2 = \sigma_3$, their product is $\det A = 15$ is the determinant of A . (in this respect singular values are like eigen values).

The columns of V are orthogonal and columns of U are orthogonal. Those columns are unit vectors after the divisions by $\sqrt{2}$ and $\sqrt{10}$, so V and U are orthogonal matrices.

$$V^T = V^{-1} \text{ and } U^T = U^{-1}$$

That orthogonality allows us to go from $AV = U\Sigma$ to the usual and famous expression of the SVP. Multiple both sides of $AV = U\Sigma$ by $V^{-1} = V^T$ the singular value decomposition of A is

$$\boxed{A = U\Sigma V^T}$$

① Find the matrices $U \in V$ for $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ with rank of A is 2. A has two positive singular values σ_1 and σ_2 .

$$\det \begin{vmatrix} 3-\lambda & 0 \\ 4 & 5-\lambda \end{vmatrix} = (3-\lambda)(5-\lambda) = 0$$

$\lambda_{\max} = 5, \lambda_{\min} = 3$

$$A^T A = \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

trace $\{A^T A\} = \text{trace } \{AA^T\} = 50$
 $\sigma_1^2 = 45 \text{ and } \sigma_2^2 = 5$

$$\sigma_1 = \sqrt{45}, \sigma_2 = \sqrt{5}, \sigma_1 \sigma_2 = 15$$

eigen values are 45 and 5.

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 45 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Then v_1 and v_2 are those orthogonal eigen vectors rescaled to length 1, divided by $\sqrt{2}$.

Right singular vectors

$$u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

left singular vectors

$$v_i = \frac{Av_i}{\sigma_i}$$

now we compute Av_1 and Av_2 which will be $\sigma_1 u_1 = \sqrt{45} u_1$

$$\sigma_2 u_2 = \sqrt{5} u_2$$

$$Av_1 = \frac{3}{\sqrt{2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sqrt{45} \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sigma_1 u_1$$

$$Av_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sqrt{5} \frac{1}{\sqrt{10}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sigma_2 u_2$$

The division by $\sqrt{10}$ makes u_1 and u_2 orthogonal. Then $\sigma_1 = \sqrt{45}$ and $\sigma_2 = \sqrt{5}$ as expected. The singular value decomposition of A is U times Σ times V^T .

$$U = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix}$$

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

U and V contains orthogonal basis for the column space and the row space of A (both spaces are just \mathbb{R}^2).

The great achievement is that those bases diagonalise A . AV equals $U\Sigma$, the matrix $A = U\Sigma V^T$ splits into two rank one matrices

columns times rows, with $\sqrt{2}\sqrt{10} = \sqrt{20}$

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \frac{\sqrt{45}}{\sqrt{20}} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{20}} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = A$$

every matrix is a sum of rank one matrices with orthogonal U 's and orthogonal V 's

Analysis of Variance [ANOVA]

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---------|------|------|----|------|----|----|
| Model A | 17.6 | 17.2 | 17 | 16.9 | 19 | 21 |
| Model B | 22 | 22.5 | 28 | 17 | 16 | 12 |
| Model C | 30.5 | 22 | 16 | 19 | 18 | 27 |

Milage of cars three models and six cars from each model.

→ one way (factor) anova.

In general, one way factor anova techniques can be used to study the effect of $k (> 2)$ levels at single factor.

→ To determine if different levels of the factor affect measured observations different.

The following hypothesis is tested.

$$H_0: \mu_i = \mu \text{ all } i=1, 2, \dots, k$$

$$H_1: \mu_i \neq \mu \text{ some } i=1, 2, \dots, k$$

where μ_i is population mean for level i .

Assumptions

- ① The observations are obtained independently and randomly from the populations defined by the factor levels.
- ② The populations at each factor level is normally distributed.
- ③ The normal populations have a common variance σ^2 . Thus for factor W

The table below shows three different brand night performance.

Life times under controlled conditions in hours excess of 1000 hours, of samples of 60 W.

Brand

| 1 | 2 | 3 |
|----|----|----|
| 16 | 18 | 26 |
| 15 | 22 | 31 |
| 13 | 20 | 24 |
| 21 | 16 | 30 |
| 15 | 24 | 24 |

| Sample size | 1 | 2 | 3 |
|-------------|------|------|------|
| sum | 80 | 100 | 135 |
| sum of sq | 1316 | 2040 | 3684 |
| mean | 16 | 20 | 27 |
| variance | 9 | 10 | 11 |

Since each of these three samples variance is an estimate of the common population variance σ^2 a pooled estimate may be calculated in usual way as follows.

$$\hat{\sigma}_w^2 = \frac{(5-1) \times 9 + (5-1) \times 10 + (5-1) \times 11}{5+5+5-3} = 10$$

This quantity is called variance within samples.

It is an estimate of σ^2 based on $V = 5+5+5-3 = 12$ degrees of freedom.

This is irrespective of whether or not the null hypothesis is true, since, differences between levels (brands) will have no effect on the within sample variance.

The variability between samples may be estimated from the three sample means

| | Brand | | |
|---------------|-------|------|----|
| | 1 | 2 | 3 |
| sample mean | 16 | 20 | 27 |
| sum | | 63 | |
| sum of sq^2 | | 1385 | |
| mean | | 21 | " |
| variance | | 31 | |

The variance divisor ($n-1$) denoted by

σ_B^2 is called the variance between

sample means. It is an estimate of

$$\frac{\sigma^2}{S} \text{ (that is } \frac{\sigma^2}{n} \text{ in general)}$$

Based upon $(3-1)=2$ degrees of freedom

but only in the null hypothesis is

true. If H_0 is false, then the

subsequent 'large' differences between
sample means will result in $S\sigma_B^2$

being inflated estimate of σ^2 .

The two estimates of σ^2 , σ_w^2 and $S\sigma_B^2$

may be tested for equality using the

F-test.

$$F = \frac{S\sigma_B^2}{\frac{\sigma_w^2}{n}}$$