Name - Vijay Gurung

ID Number - MST03 - 0070

Date - 18 - July - 2024

Topic: Mathematical Concept of Machine learning (ML)

Q:1. Formula of sigmoid function along with range?

Answer

Sigmoid Function:-

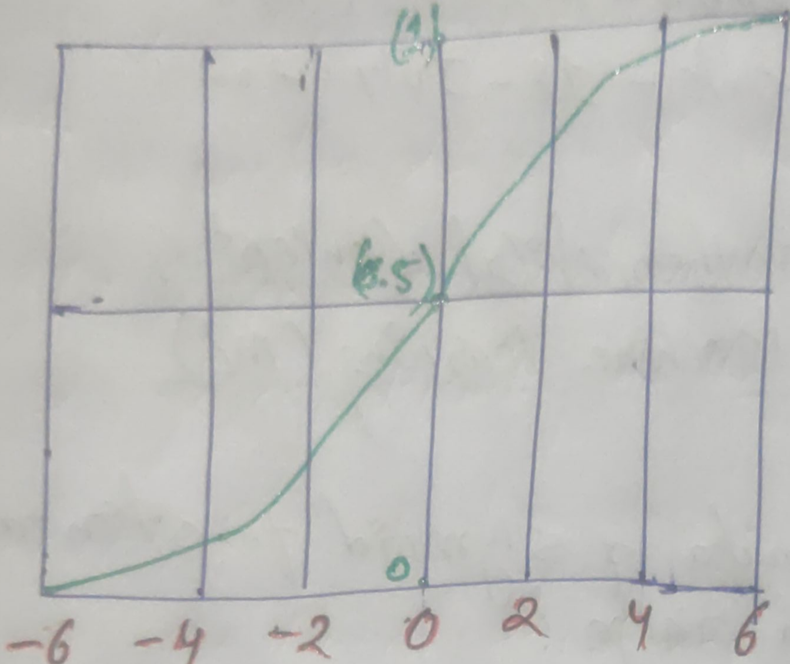
A sigmoid function ~~fn~~ is a mathematical function that produces an S-shaped curve, often used in machine learning, particularly in logistic regression and neural networks.

It is defined by the following

Formula:

$$\boxed{\sigma(x) = \frac{1}{1+e^{-x}}}$$

* Graph of Sigmoid Function :



* Range :

The output of the sigmoid function ranges from 0 to 1.

* S-shape :-

The function produces an S-shaped curve, which makes it useful for binary classification tasks.

Q.12. Explain the formula of linear kernel along with range?

Answer :- Linear Kernel function:

A linear is a type of kernel function used in support vector machines (SVMs) & other kernelized models.

It is one of the simplest types of kernel functions, used primarily for linearly seperable data.

* Formula :

$$\boxed{K(x, y) = x \cdot y}$$

* 'x' and 'y' are input vectors.
* (·) denotes the dot product (or inner product) of the vectors.

* **Range :-**

A range of the linear kernel depends on the magnitudes of the input vectors.

For vectors x & y in R^n, the range of $K(x, y)$ can be any real number, from $-\infty$ to ∞, depending on the dot product.

* **Linearity :-**

The linear kernel assumes that the relationship between the features is linear, making it a suitable choice for linearly separable data.

* **Example :-**

* **Support vector Machines (SVM) :-**

In SVMs, the linear kernel is used to find a hyperplane that best seperates the data into different classes when the data is linearly seperable.

* Text classification:

A linear kernel is often used in text classification tasks where high dimensional data (such as TF-IDF) vectors is common & linear separability is assumed.

Q:- 3. Explain the Formula use in naive Bayes classifier?

Answer

Naive Bayes classifier:

A Naive Bayes classifier is a probabilistic Machine learning model used for classification tasks.
It is based on Bayes' theorem and assumes that the features are conditionally independent given the class.

* Formula

$$\boxed{P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}}$$

* $P(A/B)$ is the Conditional probability of event (A) occurring given that event (B) has occured.

* $P(B/A)$ is the Conditional probability of event (B) occurring given that event (A) has occured.

* $P(A)$ and $P(B)$ are the independent probabilities of events (A) and (B) occuring, respectively.

In the Context of the Naive Bayes classification, the formula is used to calculate the probability of a data instance belonging to a particular class, given the features value of that instance

* Formula :

$$\boxed{P(Y=y/X=x) = \frac{P(X=x|Y=y)\,P(Y=y)}{P(X=x)}}$$

* 'Y' is the class variable
* 'X' is the feature vector
* 'y' is a particular class value
* 'x' is the feature value vector.

* **Key assumptions made by Naive Bayes classifier are as below:**

* (a) The features are independent given the class. This means the value of a features does not depend on the value of any other features, given the class.

* (b) All features contribute equally to the prediction, regardless of their type & scale.

Q!4 Explain the formula of polynomial kernel?

Answer

polynomial kernel:

The polynomial kernel is a type of kernel function used in kernel methods like support vector machines (SVMs) & Gaussian processes.

~~Per~~ Formula:

$$\boxed{K(x, y) = (x^T y + c)^d}$$

* 'x' & 'y' are input vectors
* 'c' is a constant term, often set to 1.
* 'd' is the degree of the polynomial.

A polynomial kernel maps the input vectors into a higher dimensional space where the dot product is computed.

the degree $\boxed{'d'}$ controls the flexibility

of the resulting function.
Higher degrees allow for more Complex decision boundaries.

Q: 5. Explain formula of Tan h Kernel descent?

Answer A Kernel that describes the evolution of deep artificial neural networks during training by gradient descent.

Formula: Tanh (Hyperbolic Tangent) Kernel

$$\boxed{K(x, y) = \tanh(c + x^T y)}$$

* 'x' & 'y' are the input vectors
* 'c' is a constant, often set to 1.

A Tanh kernel is a valid kernel function that maps the input vectors into a higher dimensional space where the dot product is computed.

The hyperbolic tangent function tanh(z) is defined as:

$$\boxed{\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}}$$

Q:16 Explain formula of gradient descent?

Answer <u>Gradient Descent:</u>

A Gradient descent is an optimization algorithm used to find the minimizing an ~~objects~~ objective function.

* <u>Formula:</u>

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \nabla J(\theta^{(t)})$$

* theta (θ)^(t) is the parameter vector at iteration 't'.
* Alpha is the learning rate also called the step size).
* nabla J(theta(θ)^(t)) is the gradient of the objective function (J) with respect to the parameters theta (θ) at iteration $\boxed{t}$.

Q! 7 Explain all three formula of linear regression metrics ?

Answer

The three formula of linear regression metrics are as follows:-

(a) Mean Squared Error (MSE)

MSE measures the average squared difference between the actual and predicted values. It penalizes large errors more heavily.

* Formula :-

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

* 'n' is the number of data points
* 'y_i' is the actual value
* 'hat(y)_i' is the predicted value

(b) R- Squared (R^2) :

R-squared measures the proportion of variance in the target variable that is predictable from the independent variables. It ranges from 0 to 1., with higher values indicating better model fit.

* Formula :

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

* n' is the number of data points
* 'y_i' is the actual value
* 'hat(y)-i' is the predicted value
* bar (y) - is the mean of the actual values.

(c) **Root Mean Squared Error (RMSE):**

RMSE is the square root of the MSE. It is in the same units as the target variable, making it easier to interpret.

* **Formula:**

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

* 'n' is the number of data points
* 'y_i' is the actual value
* 'hat y_i' is the predicted value.

Q: 8 Explain the formula of cost function?

Answer Cost Function:

A cost function is a mathematical formula that calculate the total cost of ~~p~~ production given a specific amount of goods produced.

It shows how costs will change at different output levels.

* Formula:

$$\boxed{C(x) = F + Vx}$$

* $C(x)$ is the total cost ~~of~~ to product 'x' units of a good.
* 'F' is the fixed cost of production (Costs that do not change with output).

* ~~V~~

* 'V' is the variable cost per unit of a good.

* 'x' is the total number of goods produced.

*(1) DESCRIPTIVE STATISTICS

(A) CENTERAL TENDENCY

MEAN | MEDIAN | MODE

(1) MEAN :-

MEAN is the average of a set of numbers. It is calculated by adding up all the numbers and dividing by the total count of numbers.

Formula :-

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where,

* $\bar{x}$ It represents the mean.
* $\sum_{i=1}^{n} x_i$ represents the sum of all the values in the set.
* 'n' represents the total number of values in the set.

Example :-

In the set [5, 10, 15, 20], the mean is calculated as :

$$\text{Mean} = \frac{5+10+15+20}{4} = \frac{50}{4}$$
$$= 12.5$$

(2) MEDIAN :-

The MEDIAN is the middle value in a sorted set of numbers: If there are an odd number of values, the median is the middle value.
If there are an even number of values, the median is the average of the two middle values.

Formula :-

(a) For an odd number of values.

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{value}$$

(b) For an even number of values:

$$\text{Median} = \left(\frac{n}{2}\right)^{th} \text{value} + \left(\frac{n}{2}+1\right)^{th} \text{value}$$

* 'n' It represent the total number of values in the set.

Example :-

* In the set [5, 10, 15, 20, 25].
The ~~me~~ middle value $\boxed{15}$ is the median.

* In the set [5, 10, 15, 20]
The median is -

$$\text{Median} = \frac{10+15}{2}$$
$$= 12.5$$

(3) MODE :-

In Mode the value that appears most frequently in a set of numbers.

Example :-

(a) In the set [5, 10, 15, 20, 10, 5]
The mode is 5 & 10, why because both appear twice, which is more than any other number.

(b) In the set [5, 10, 15, 20], there is no mode as each number appears only once.

Summary :-

(1) Mean :
The average of a set of numbers

(2) Median :
The middle value in a sorted set of numbers.

(3) mode :
The value that appears most frequently in a set of numbers.

(B) MEASURES of VARIABILITY (spread)

These measures indicate the ~~spead~~ spread or dispersion of the dataset.

(1) Range (2) variance (3) standard deviation

(1) Range :
The difference between the highest and lowest values.

(2) variance :
The average of the squared differences from the mean.

(3) standard deviation :
The square root of the variance, indicating how much the values in the dataset deviate from the mean.

Example :-

~~Consider~~

Consider the data set [10, 20, 20, 30, 40]

(1) Range :

$$\text{Range} = 40 - 10$$
$$= 30$$

(2) variance :

(i) Calculate the mean : 24

(ii) Then, find the squared deviations and their average :

$$(10-24)^2 = 196$$
$$(20-24)^2 = 16$$
$$(20-24)^2 = 16$$
$$(30-24)^2 = 36$$
$$(40-24)^2 = 256$$

(iii) Sum of square deviations :

$= 196 + 16 + 16 + 36 + 256$

$= 520$

(iv) variance $= \dfrac{520}{5}$

$= 104$,,

(3) standard deviation :

standard deviation $= \sqrt{104} \approx 10.2$,,

* OTHER DESCRIPTIVE STATISTICS

(1) Minimum and Maximum :

The smallest and largest values in the dataset.

(2) Quartiles :

values that divide the dataset into four equal parts.

(3) Interquartile Range (IQR) :

The range of the middle 50% of the data, calculated as Q3 - Q1.

* descriptive statistics provide a comprehensive summary of the dataset, allowing for a better understanding of the distribution, central value & spread of the ~~data~~. data.

Example :

Consider the dataset : [10, 20, 20, 30, 40]

* Minimum : 10

* Maximum : 40

* Quartiles :

    * Q1 (25th percentile) : 20
    * Q2 (50th percentile or median) : 20
    * Q3 (75th percentile) : 30

* Interquartile Range (IQR) :

$$IQR = Q3 - Q1 = 30 - 20$$
$$= 10$$

Conclusion

* Central Tendency : Mean (24), Median (20), Mode (20)

* variability : Range (30), variance (104), standard deviation (≈ 10.2)

* other statistics : Minimum (10), Maximum (40), Quartiles (20, 20, 30), IQR 10.

STATISTICS

(2) INFERENTIAL STATISTICS

Inferential statistics involves making inferences about populations using data drawn from the population.
Instead of merely describing the data as ~~an~~ in descriptive statistics, inferential statistics allow us to make predictions or generalizations about a larger group based on a sample of data.

* Key Concepts in Inferential statistis

(1) population vs. sample
(2) parameter vs. statistic
(3) Hypothesis Testing.
(4) Confidence Intervals
(5) Regression Analysis
(6) T-Tests and ANOVA

(1) population vs. sample.

* population :

The entire group that we want to draw conclusions about.

* sample :

A subset of the population used to collect data and make inferences about the population.

(2) parameter vs. statistic :

* parameter :-

A numerical characteristic of a population (eg, population mean).

* statistic :

A numerical characteristic of a sample (eg, sample mean).

(3) Hypothesis Testing:

* **Null Hypothesis (H_0):**
The hypothesis that there is no effect or no difference.

* **Alternative Hypothesis (H_a):**
The hypothesis that there is an effect or a difference.

* **p-value:**
The probability of observing the data, or something more extreme, if the null ~~hypo~~ hypothesis is true.

* **Significance level (α):**
A threshold for determining wether the p-value is low enough to reject the null hypothesis (commonly 0.05).

(4) Confidence Intervals:

* A range of values that is likely to contain the population parameter with a certain level of confidence (e.g, 95%).

(5) Regression Analysis:

A statistical method for examining the relationship between two or more variables.

(6) T-tests and ANOVA:

* T-Test: used to compare the means of two groups.

* ANOVA (Analysis of variance):
used to compare the means of three or more groups.

* EXAMPLE OF INFERENTIAL STATISTICS

Q: Scenario:-

A Company wants to know the average amount of time its employees spend on breaks during a work days.

Instead of asking all 1,000 employees, the Company surveys a sample of 100 employees.

(1) Sample data Collection:

* Suppose the average break time from the sample of 100 employees is ~~30 minutes~~ 30 minutes with a standard deviation of 5 minutes.

(2) Estimate Population Mean:

* The Company uses the sample mean to estimate the population mean.

(3) Confidence Interval:

* Calculate the 95% Confidence Interval for the population mean break time.

* Formula:- $CI = \bar{x} \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$

* where,

  * $\bar{x}$ is the sample mean (30 minutes)
  * z is the ~~zero~~ z-score corresponding level (1.96 for 95%).
  * σ is the sample standard deviation (5 minutes)
  * n is the sample size (100)

* Calculations:

$$CI = 30 \pm 1.96\left(\frac{5}{\sqrt{100}}\right)$$

$$CI = 30 \pm 1.96\,(0.5)$$

$$CI = 30 \pm 0.98$$

$$CI = [29.02, 30.98]$$

* Interpretation:

we are 95% Confident that the true mean break time for all employees is between 29.02 and 30.98 minutes.

(4) Hypothesis Testing:

* Suppose the Company has a policy that States the average break time should be 28 minutes.

* Null Hypothesis (H_0): The average break time is 28 minutes.

* Alternative Hypothesis (H_a): The average break time is not 28 minutes.

* Calculate the test statistic (t-Score):

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{30 - 28}{\frac{5}{\sqrt{100}}}$$

$$t = \frac{2}{0.5}$$

$$t = 4$$

* Compare the t-score to the critical value from the t-distribution table with $\boxed{df = n-1 = 99}$

* If the t-score exceeds the critical value, reject the null hypothesis.

(5) Conclusion :

If the p-value corresponding to the t-score is less than 0.05, we reject the null hypothesis and conclude that the average break time is significantly different from ~~28 min~~ 28 minutes.

Summary

* Inferential statistics allows us to make predictions or generalizations about a populations based on a sample.

* Key techniques include hypothesis testing, Confidence intervals and regression analysis.

* Example:
Estimating the average break time for all employee based on a sample and testing if it differs from a specified value.

~~Safe~~.
Inferential statistics provide powerful tools to make decisions and predictions about populations based on sample data, allowing for informed decision-making in various fields.