What is Hypothesis?

A Guess,Assupmtion,Intuition, Feeling, Instinct,

(In Hindi)Anuman,Atkal,Kalpana.

A concept that is not yet verified but that if true would

explain certain facts.

//Concept Learning

Inferring a boolean-valued function from training examples

of its input and output.

Acquiring the definition of a general category from

given sample positive and negative training examples of the category.

Concept learning can be viewed as the task of searching through a large

space of hypotheses implicitly defined by the hypothesis

representation.

The goal of this search is to find the hypothesis that best fits the

training examples.

H : Hypothesis Space.

h: A hypothesis h in H.

X : Instance Space.

x : An instance from instance space X

D : The Training Data

h(x) : The prediction produced by hypothesis h for instance x.

c(x) : Class / Target Label.

There are 96 (= 3.2.2.2.2.2) distinct instances in X.

There are 5120 (=5.4.4.4.4.4) syntactically distinct hypotheses in H.

Two more values for attributes: ? and Ø

Every hypothesis containing one or more Ø symbols represents the

empty set of instances; that is, it classifies every instance as

negative.

There are 973 (= 1 + 4.3.3.3.3.3) semantically distinct hypotheses in H.

Only one more value for attributes: ?, and one hypothesis representing empty set of

instances.

//FIND S ALGORITHM?

FIND-S Algorithm starts from the most specific hypothesis and generalize it by considering only positive examples.

FIND-S algorithm ignores negative examples.

FIND-S algorithm finds the most specific hypothesis within H that is consistent with the positive training examples.

//LIMITATION OF FIND S

* There is no way to determine if the hypothesis is consistent throughout the data.
* Inconsistent training sets can actually mislead the Find-S algorithm, since it ignores the negative examples.
* A good concept learning algorithm should be able to backtrack the choice of hypothesis found so that the resulting hypothesis can be improved over time. Unfortunately, Find-S provide no such method.

//CANDIDATE ELIMINATION ALGORITHM

The key idea in the Candidate-Elimination algorithm is to output a description of the set of all hypotheses consistent with the training examples.

The key difference between this definition of **consistent** and **satisfies.**

An example x is said to satisfy hypothesis h when **h(x) = 1**, regardless of whether x is a positive or negative example of the target concept.

However, whether such an example is consistent with h depends on the target concept, and in particular, whether **h(x) = c(x).**

**//TYPES of ANN**

1. Single Layered FEED Forward (Perceptron Learning)
2. Multi Layered Feed Forward
3. Recurrent

//Perceptron Learning

The perceptron is used for binary classification.

The perceptron can only model linearly separable classes.

**Bias** : move the hyperplane which separate two classes from the origin.

**Learning rate** : moderate the degree to which the weight change during the training process.

**Activation function :** Introduce Non-Linearity, Feature Extraction:

//Recurrent Neural Network(RNN)

The principle of Recurrent Neural Network is to feedback the output of a layer back to the input again. This principle helps to predict the outcome of the layer. In the Computation process, Each neuron will act as a memory cell. The neuron will retain some information as it goes to the next time step.

It is called a recurrent neural network process. The data to be used later will be remembered and work for the next step will go on in the process. The prediction will improve by error correction. In error correction, some changes are made to create the right prediction output. The learning rate is the rate of how fast the network can make the correct prediction from the wrong prediction.

There is much application of Recurrent Neural Networks, and one of them is the model of converting text to speech. The recurrent neural network was designed for supervised learning without any requirement of teaching signal.

//BACK PROPAGAGTION ALGO

Here's an overview of the backpropagation algorithm:

Forward Pass: Input data is fed forward through the network. The network computes an output based on the current weights and biases using an activation function.

Error Computation: The error is computed by comparing the network's output with the desired output (the target).

Backward Pass: The error is propagated backward through the network. This involves computing the gradient of the error with respect to the weights and biases at each layer of the network using the chain rule from calculus.

Weight Update: The weights and biases are then adjusted to minimize the error. This is done by moving in the opposite direction of the gradient with a certain step size determined by the learning rate. The learning rate determines how big a step is taken during each iteration.

//Accuracy

**Precision — *What percent of your predictions were correct?***

Precision is the ability of a classifier not to label an instance positive that is actually negative. For each class, it is defined as the ratio of true positives to the sum of a true positive and false positive.

Precision:- Accuracy of positive predictions.

Precision = TP/(TP + FP)

**Recall — *What percent of the positive cases did you catch?***

Recall is the ability of a classifier to find all positive instances. For each class it is defined as the ratio of true positives to the sum of true positives and false negatives.

Recall:- Fraction of positives that were correctly identified.

Recall = TP/(TP+FN)

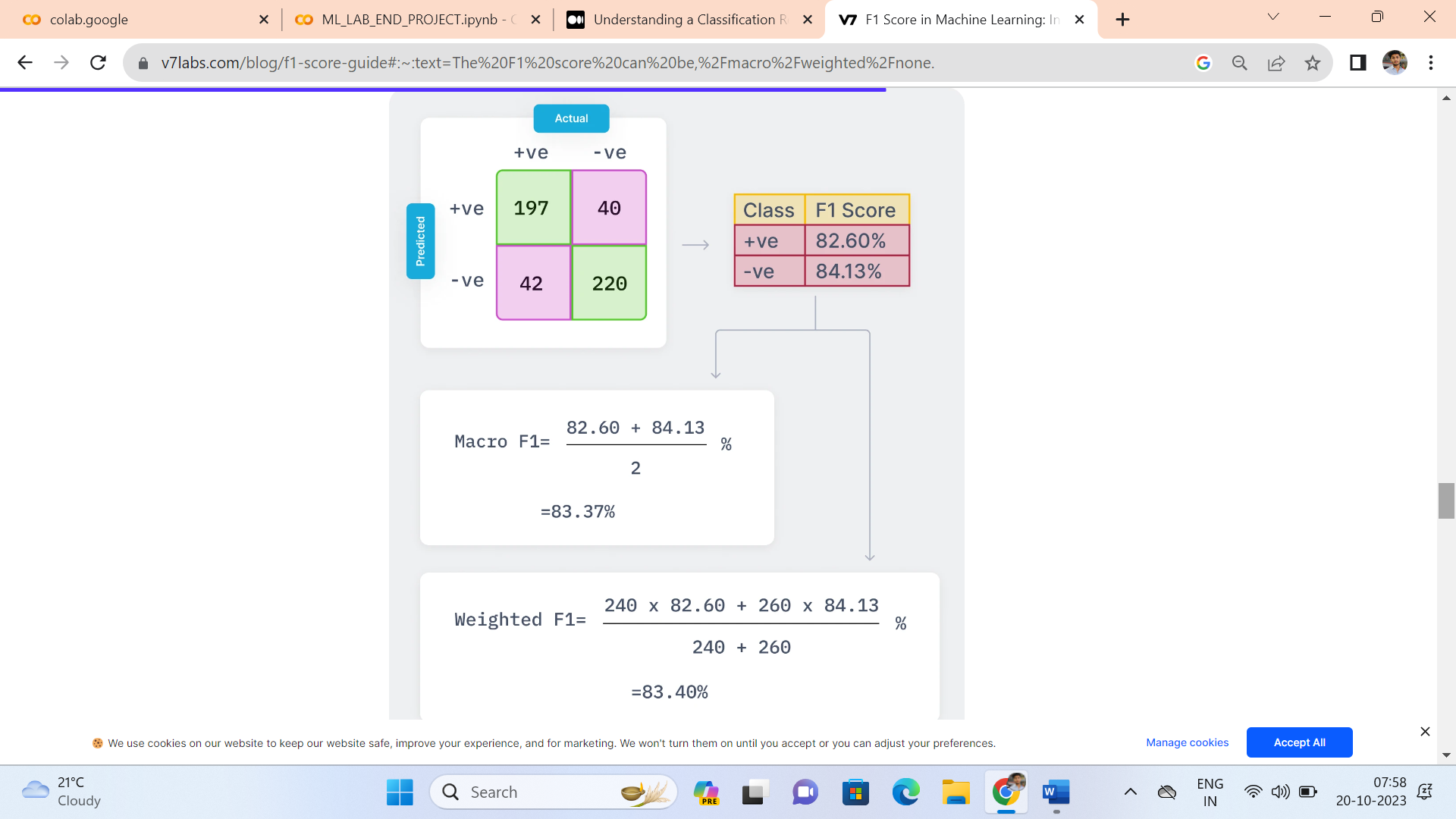
**F1 score — *What percent of positive predictions were correct?***

The F1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0. F1 scores are lower than accuracy measures as they embed precision and recall into their computation. As a rule of thumb, the weighted average of F1 should be used to compare classifier models, not global accuracy.

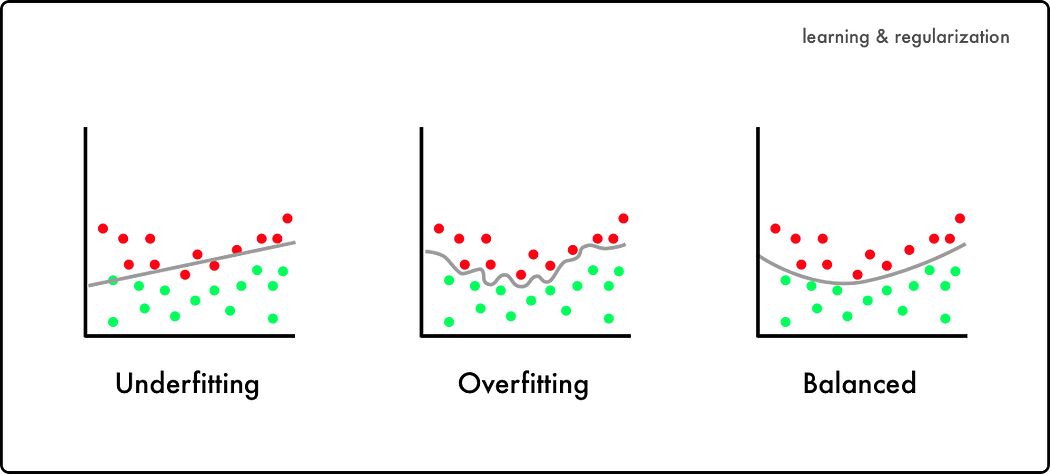
F1 Score = 2\*(Recall \* Precision) / (Recall + Precision)

**Support**

Support is the number of actual occurrences of the class in the specified dataset. Imbalanced support in the training data may indicate structural weaknesses in the reported scores of the classifier and could indicate the need for stratified sampling or rebalancing. Support doesn’t change between models but instead diagnoses the evaluation process.



//OVERFITTING UNDERFITTING and BALANCED



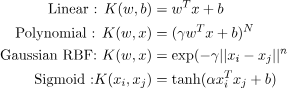
**1) Support Vector Machine(SVM)**

* Support Vector Machine (SVM) is a powerful machine learning algorithm used for linear or nonlinear classification, regression, and even outlier detection tasks. (it’s best suited for classification.)
* The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space
* he hyperplane tries that the margin between the closest points of different classes should be as maximum as possible.
* The dimension of the hyperplane depends upon the number of features.
* **hard margin SVM** aims to find a hyperplane that perfectly separates the data into two classes without any misclassification, while **soft margin SVM** allows some misclassification by introducing slack variables

Slack variables can be added to allow misclassification of difficult or noisy examples.

***Terminology of SVM.***

* **Hyperplane:** Hyperplane is the decision boundary that is used to separate the data points of different classes in a feature space. In the case of linear classifications, it will be a linear equation i.e. wx+b = 0.
* **Support Vectors**: Support vectors are the closest data points to the hyperplane, which makes a critical role in deciding the hyperplane and margin.
* **Margin:** Margin is the distance between the support vector and hyperplane. The main objective of the support vector machine algorithm is to maximize the margin. The wider margin indicates better classification performance.
* **Kernel:** Kernel is the mathematical function, which is used in SVM to map the original input data points into high-dimensional feature spaces, so, that the hyperplane can be easily found out even if the data points are not linearly separable in the original input space. Some of the common kernel functions are linear, polynomial, radial basis function(RBF), and sigmoid.



* **Hard Margin:** The maximum-margin hyperplane or the hard margin hyperplane is a hyperplane that properly separates the data points of different categories without any misclassifications.
* **Soft Margin:** When the data is not perfectly separable or contains outliers, SVM permits a soft margin technique. Each data point has a slack variable introduced by the soft-margin SVM formulation, which softens the strict margin requirement and permits certain misclassifications or violations. It discovers a compromise between increasing the margin and reducing violations

***Advantage :***

* Effective in high-dimensional cases.
* Its memory is efficient as it uses a subset of training points in the decision function called support vectors.
* Different kernel functions can be specified for the decision functions and its possible to specify custom kernels.

**Disadvantage: does not perform well when dataset is large**

* The biggest limitation of Support Vector Machine is the choice of the kernel. The wrong choice of the kernel can lead to an increase in error percentage.
* With a greater number of samples, it starts giving poor performances.
* SVMs have good generalization performance but they can be extremely slow in the test phase.
* SVMs have high algorithmic complexity and extensive memory requirements due to the use of
* quadratic programming.