Difference between Multilabel and Multiclass classification

Normalization vs Standardization (when to use what)

What is tfidf

Bias(underfitting) vs variance(overfitting)

**Fixing High Bias:**

Obtaining new features

Fitting a more complex function

Adding more layers in neural network

Increasing the number of more input parameters flowing into the model in neural network

**Fixing High Variance:**

Getting more training examples

Taking smaller set of features

Adding more training samples in neural networks

Online Learning

Confusion Matrix

Precision vs Recall

When you plot a learning curve(training error,cross validation error as a function of number of training examples) if training error and cross validation error both are high then high bias or high variance **High Bias**

If you train a model and it performs well on the training set and fails to generalize on the validation set then High bias or high variance **High Variance**

Adding new features fixes High Bias or High Variance **High Bias**

In case of neural network adding more layers reduces bias or variance **Bias**

What is Online Learning(Normal gradient descent vs stochastic gradient descent)

Difference between Multilabel and Multiclass classification

What is tfidf tf\*idf

tf=number of times the word appears in a document

Idf = log(total number of documents/number of documents in which this word appears)

PageRank Algorithm and its applications(Word Sense Disambiguation what specific meaning is being conveyed by the given sentence whenever it’s appearing, Auto Summarization of Text(TextRank))

How would you build a model to distinguish between apple company and apple fruit

Standardization vs Normalization when to use what

Standardization: mean of 0 and standard deviation 1

60% of the value lies in between -1 and 1

Normalization : num-min/max-min 0-1 range

Doesnt decrease the importance of outliers

Min- Max tries to get the values closer to mean. But when there are outliers in the data which are important and we don’t want to loose their impact ,we go with Z score normalization.

If your data contains more than 30% missing values what would you do ??

-> Perform treatement or drop them from the analysis

Why do we prefer one hot encoding over label encoding

Accuracy of a model is 99. Do you thing the model is good to go with.

Feature is 70% sparse how would you handle the data

When do you take log of a feature

Skewed classes dealing

What is the role of bias in a neural network

The bias value allows the activation function to be shifted to the left or right, to better fit the data. Hence changes to the weights alter the steepness of the sigmoid curve, whilst the bias offsets it, shifting the entire curve so it fits better. Note also how the bias only influences the output values, it doesn’t interact with the actual input data

## **New Questions**

**1.Explain Stochastic Gradient Descent to a five year old.**

**2.What is dropout? What happens if we apply it to a neural network ?**

**3.How do we use a 1x1 convolution unit ? (**[**https://www.youtube.com/watch?v=vcp0XvDAX68**](https://www.youtube.com/watch?v=vcp0XvDAX68)**)**

**4.When would you use ML? When would you use NN?**

**5.What is dimensionality reduction ? How is it useful ?**

The goal is to simplify data without losing too much information. It is important because a. It helps to run the algorithm much faster b. The data will take less memory and space

**6.Why is logistic regression called regression when we are using it for classification instead ?**

**7. What are some challenges faced in machine learning ?**

There are multiple challenges that can be categorised due to two main causes

1. Bad Data
2. Insufficient quantity of training data
3. Non representative Training Data
4. Poor Quality of Data
5. Irrelevant Features

**Solution : Feature Engineering :** Selection, Extraction and Creation of better features

1. Bad Algorithm
   1. Underfitting the training data
   2. Overfitting the training data

**Solution**

|  |  |
| --- | --- |
| **Underfitting** | **OverFitting** |
| 1.Using a powerful model | 1.Reducing the no. parameters |
| 2.Finding better features | 2.Get more training data |
| 3.Reducing regularization hyperparameters | 3.Reduce noise in training data |
|  | 4.Using regularization |

**8.What is validation set used for?**

**9.What is the no free lunch theorem ?**

This implies that a model that explains a certain situation well may fail in another situation. In both statistics and machine learning, we need to check our assumptions before relying on a model. The “No Free Lunch” theorem states that there is no one model that works best for every problem

**10. Why is MSE suited for Linear Regression as a cost function ?**

The MSE cost function for a Linear Regression model happens to be a convex function, which means that if you pick any two points on the curve, the line segment joining them never crosses the curve. This implies that there are no local minima, just one global minimum.

It is also a continuous function with a slope that never changes abruptly.

These two facts have a great consequence: Gradient Descent is guaranteed to approach arbitrarily close the global minimum (if you wait long enough and if the learning rate is not too high

**11. Does Stochastic Gradient Descent always return optimal parameter values. If not, why ?**

Due to its stochastic (i.e., random) nature, this algorithm is much less regular than Batch Gradient Descent: instead of gently decreasing until it reaches the minimum, the cost function will bounce up and down, decreasing only on average.

Over time it will end up very close to the minimum, but once it gets there it will continue to bounce around, never settling down . So once the algorithm stops, the final parameter values are good, but not optimal.

**12. How do you tackle the randomness in Stochastic Gradient Descent ?**

Randomness is good to escape from local optima, but bad because it means that the algorithm can never settle at the minimum. One solution to this dilemma is to gradually reduce the learning rate. The steps start out large (which helps make quick progress and escape local minima), then get smaller and smaller, allowing the algorithm to settle at the global minimum. This process is called simulated annealing, because it resembles the process of annealing in metallurgy where molten metal is slowly cooled down.

The function that determines the learning rate at each iteration is called the learning schedule. If the learning rate is reduced too quickly, you may get stuck in a local minimum, or even end up frozen halfway to the minimum. If the learning rate is reduced too slowly, you may jump around the minimum for a long time and end up with a suboptimal solution if you halt training too early

**COMPUTER VISION AND SELF DRIVING CARS**

1. **What is computer vision ?**

Computer Vision is a subset of AI where we try to teach computers to gain high level information from images and video

1. **Why is C++ preferred for implementing computer vision algorithms**

Speed: C++ beats other languages and takes less time and memory

Community and Libraries support

Portability: easier to port C++ into code bases of other languages but not the other way round

1. **Applications of Computer Vision**

Object Detection

Object Recognition

Object Localization

1. **Commonly used Algorithms in Self Driving Cars**

AdaBoost: Adaptive Boosting

K-means clustering

Principal Component Analysis [PCA]

Histogram of Gradients [HOG]

Support Vector Machines [SVM]

Neural Network Regression

1. **Commonly used Mathematical Methods used in Computer Vision**

According to this course: [CS491Y/791Y Mathematical Methods for Computer Vision](http://www.cse.unr.edu/~bebis/MathMethods/)

They are:

1. Linear Algebra
2. Singular Value Decomposition
3. Introductory level Pattern Recognition
4. Principal Component Analysis
5. Linear Discriminant Analysis
6. Fourier Transform
7. Wavelets
8. Probability, Bayes rule, Maximum Likelihood, MAP
9. Mixtures and Expectation-Maximization Algorithm
10. Introductory level Statistical Learning
11. Support Vector Machines
12. Genetic Algorithms
13. Hidden Markov Models
14. Bayesian Networks
15. Kalman filtering
16. **How many channels are there in Grayscale Image and RGB image ?**

**LINEAR ALGEBRA**

1. **What is broadcasting in connection to Linear Algebra?**
2. **What are scalars, vectors, matrices, and tensors?**
3. **What is Hadamard product of two matrices?**
4. **What is an inverse matrix?**
5. **If inverse of a matrix exists, how to calculate it?**
6. **What is the determinant of a square matrix? How is it calculated? What is the connection of determinant to eigenvalues?**
7. **Discuss span and linear dependence.**
8. **What is Ax = b? When does Ax =b has a unique solution?**
9. **In Ax = b, what happens when A is fat or tall?**
10. **When does inverse of A exist?**
11. **What is a norm? What is L1, L2 and L infinity norm?**
12. **What are the conditions a norm has to satisfy?**
13. **Why is squared of L2 norm preferred in ML than just L2 norm?**
14. **When L1 norm is preferred over L2 norm?**
15. **Can the number of nonzero elements in a vector be defined as L0 norm? If no, why?**
16. **What is Frobenius norm?**
17. **What is a diagonal matrix?**
18. **Why is multiplication by diagonal matrix computationally cheap? How is the multiplication different for square vs. non-square diagonal matrix?**
19. **At what conditions does the inverse of a diagonal matrix exist?**
20. **What is a symmetrix matrix?**
21. **What is a unit vector?**
22. **When are two vectors x and y orthogonal?**
23. **At R^n what is the maximum possible number of orthogonal vectors with non-zero norm?**
24. **When are two vectors x and y orthonormal?**
25. **What is an orthogonal matrix? Why is computationally preferred?**
26. **What is eigendecomposition, eigenvectors and eigenvalues?**
27. **How to find eigen values of a matrix?**
28. **Write the eigendecomposition formula for a matrix. If the matrix is real symmetric, how will this change?**
29. **Is the Eigendecomposition guaranteed to be unique? If not, then how do we represent it?**
30. **What are positive definite, negative definite, positive semi definite and negative semi definite matrices?**
31. **What is Singular Value Decomposition? Why do we use it? Why not just use ED?**
32. **Given a matrix A, how will you calculate its Singular Value Decomposition?**
33. **What are singular values, left singulars and right singulars?**
34. **What is the connection of Singular Value Decomposition of A with functions of A?**
35. **Why are singular values always non-negative?**
36. **What is the Moore Penrose pseudo inverse and how to calculate it?**
37. **If we do Moore Penrose pseudo inverse on Ax = b, what solution is provided is A is fat? Moreover, what solution is provided if A is tall?**
38. **Which matrices can be decomposed by ED?**
39. **Which matrices can be decomposed by SVD?**
40. **What is the trace of a matrix?**
41. **How to write Frobenius norm of a matrix A in terms of trace?**
42. **Why is trace of a multiplication of matrices invariant to cyclic permutations?**
43. **What is the trace of a scalar?**
44. **Write the frobenius norm of a matrix in terms of trace?**