https://www.analyticsvidhya.com/blog/2020/04/comprehensive-popular-deep-learning-interview-questions-answers/

Supervised Learning and Deep learning:-

* Image classification
* Text classification
* Sequence tagging

Unsupervised deep learning techniques:-

* Word embeddings (like Skip-gram and Continuous Bag of Words): Understanding Word Embeddings: From Word2Vec to Count Vectors
* Autoencoders: Learn How to Enhance a Blurred Image using an Autoencoder!

**Weights and Bias with Cricket eg :-**

**Factors** like the weather or temperature might have a higher weight, and other factors like equipment would have a lower weight.

**Bias** lets you assign some threshold which helps you activate a decision-point (or a neuron) only when that threshold is crossed.

* need 1 ball and 6 wickets

**Forward propagation:**  calculate the output of the activation at each node .

**Backpropagation :** minimize the cost function by its understanding of how it changes with changing the weights and biases in a neural network.

Change obtained by gradient descent.

**Batch Normalization** is one of the techniques used for reducing the training time of our deep learning algorithm. Just like normalizing our input helps improve our logistic regression model, we can normalize the activations of the hidden layers in our deep learning model as well

### CNN :- it just identifies curves and edges. Thus, instead of looking at the entire image, it helps to just read the image in parts.

### Element wise operations:-

### CNN_Deep_learning_questions

### Artificial Neural Network (ANN) :- each input is given the same weight and fed to the network at the same time

### I saw the movie and hated it

### Difficult to identify how it is attached to movie to RNN is comes into picture.

### RNN :- The addition of a loop is to denote preserving the previous node’s information for the next node, better for sequential data.

### Valid Padding: When we do not use any padding. The resultant matrix after convolution will have dimensions (n – f + 1) X (n – f + 1)

### Same padding: Adding padded elements all around the edges such that the output matrix will have the same dimensions as that of the input matrix.

### Gradient Descent algorithm tries to minimize the error by taking small steps towards the minimum value. These steps are used to update the weights and biases in a neural network.

### exploding gradient :- the steps become too large and this results in larger updates to weights and bias terms

### vanishing gradient : steps are too small and this leads to minimal changes in the weights and bias terms – even negligible changes at times.

### LSTM – Input gate, Output gate, forget gate

### GRU :- Update Gate, Ouput gate

### GRU less complex and faster than LSTM

### Transformers use what is called the attention mechanism. This basically means mapping dependencies between all the parts of a sentence

<https://www.edureka.co/blog/interview-questions/deep-learning-interview-questions/>

### Ridge Regression:

* **Linear regression** in which a small amount of bias is introduced
* **lambda \* the squared weight** of each individual features
* Linear or polynomial regression will **fail** if there is high **collinearity** between the independent variables

Regression Analysis in Machine learning

### Lasso Regression

* Technique to reduce the complexity of the model
* Penalty term contains only the **absolute weights instead** of a square of weights

Regression Analysis in Machine learning

### Decision Tree Regression:

* A **supervised** learning algorithm
* Decision Tree regression builds a tree-like structure
* Random forest is one of the most **powerful** supervised learning algorithms
* The Random Forest regression is an ensemble learning
* **Ensemble learning** - combines multiple decision trees and predicts the final output based on the average of each tree output

Random forest uses **Bagging and Boosting**

**Bagging – Mostly used**

* **Random set**
* **Parrallel training**
* **Aggragate all tree**

### Support Vector Regression

* **Kernel:**  function used to map a lower-dimensional data into higher dimensional data.
* **Hyperplane:** **SVM**, it is a **separation** line between two classes, but in **SVR**, it is a line which helps to **predict** the continuous.
* **Boundary line:**  the two lines apart from hyperplane, which creates a margin for datapoints.
* **Support vectors:**  the datapoints which are nearest to the hyperplane and opposite class.

**Linear Regression in Machine Learning:-**

* It is used for **predictive analysis**
* **Goal is to find the best fit line**
* **Error** between predicted values and actual values should be **minimized**
* For Linear Regression, we use the **Mean Squared Error (MSE)**

**Residuals:** The distance between the actual value and predicted values is called residual.

Linear Regression in Machine Learning

**Residuals:** The distance between the actual value and predicted values is called residual.

**Gradient Descent:**

* Used to minimize the MSE by calculating the gradient of the cost function.
* A regression model uses gradient descent to update the coefficients of the line by reducing the cost function.
* It is done by a random selection of values of coefficient and then iteratively update the values to reach the minimum cost function.

Model Performance:

The Goodness of fit determines how the line of regression fits the set of observations

**1. R-squared method:**

* R-squared is a statistical method that determines the goodness of fit.Linear Regression in Machine Learning

## **Evaluating a Classification model:**

**1. Log Loss or Cross-Entropy Loss:**

* + value between the 0 and 1.
  + lower log loss -> higher accuracy of the mode

**2. Confusion Matrix:**

**3. AUC-ROC curve**

### Feature Selection :-

* **Filtering Methods**
* **Wrapper Methods**
* **Embedded Methods**

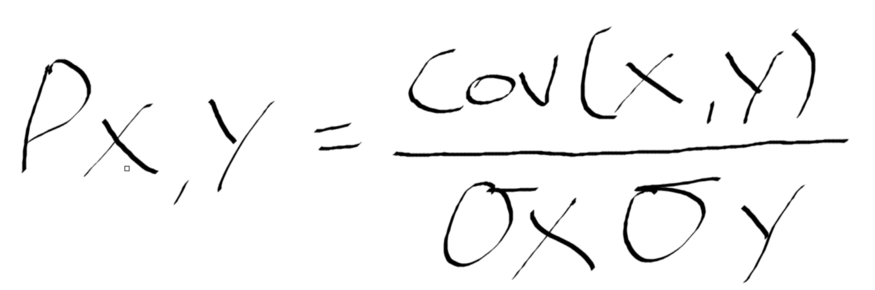
https://www.kdnuggets.com/2020/08/getting-started-feature-selection.html

**Filtering Methods**

Statistically significant relationship from each input feature(x) to the target feature(y)

An example of a filtering method is Pearson's correlation coefficient

It ranges from +1 to -1



 Cov is the covariance,

σX is the standard deviation of X,

σY is the standard deviation of Y.

1. Chi-Squared Test
2. Information Gain
3. F Test
4. ANOVA
5. Correlation Coefficient Scores (eg. Pearson’s Correlation Coefficient)

**2. Wrappers Methods**

some features are fed to the ML model, and evaluate the performance.

The performance decides whether to add those features or remove to increase the accuracy of the model.

* Forward Selection
* Backward Selection
* Bi-directional Elimination

**3. Embedded Methods:** Embedded methods check the different training iterations of the machine learning model and evaluate the importance of each feature. Some common techniques of Embedded methods are:

* **LASSO**
* **Elastic Net**
* **Ridge Regression, etc.**

**Techniques handle imbalanced data :**

<https://www.kdnuggets.com/2017/06/7-techniques-handle-imbalanced-data.html>

<https://www.kdnuggets.com/2021/07/distinguish-yourself-hundreds-other-data-science-candidates.html>

https://www.analyticsvidhya.com/blog/2017/03/imbalanced-data-classification/

### 1. Use the right evaluation metrics

* **Precision/Specificity**: how many selected instances are relevant.
* **Recall**/Sensitivity: how many relevant instances are selected.
* **F1 score:** harmonic mean of precision and recall.
* **MCC**: correlation coefficient between the observed and predicted binary classifications.
* **AUC**: relation between true-positive rate and false positive rate.

### 2. Resample the training set

**2.1. Under-sampling :**

* **R**educing the size of the abundant class.
* used when quantity of data is sufficient.

**2.2. Over-sampling**

* It tries to balance dataset by increasing the size of rare samples.
* New rare samples are generated by using e.g. **repetition, bootstrapping or SMOTE (Synthetic Minority Over-Sampling Technique)**

### 3. Use K-fold Cross-Validation in the right way

* Used over-sampling method to address imbalance problems.
* It is applied after **over-sampling**
* oversampling can lead to **overfitting**
* cross-validation should always be **done before** over-sampling the data.
* Only by resampling the data repeatedly, **randomness** can be introduced into the dataset to make sure that there won’t be an overfitting problem.

### 4. Ensemble different resampled datasets

### Happen in logistic regression or random forest tend to generalize by discarding the rare class

### Best to use all the samples of the rare class and n-differing samples of the abundant class

### you just split the 10.000 cases with 1 rare case in 10 chunks and train 10 different models.

### 7. Design your own models

### The famous XGBoost is already a good starting point if the classes are not skewed too much, because it internally takes care that the bags it trains on are not imbalanced. But then again, the data is resampled, it is just happening secretly.

### By designing a cost function that is penalizing wrong classification of the rare class more than wrong classifications of the abundant class, it is possible to design many models that naturally generalize in favour of the rare class. For example, tweaking an SVM to penalize wrong classifications of the rare class by the same ratio that this class is underrepresented.

### A discriminative model makes predictions based on conditional probability and is either used for classification or regression.

### A generative model revolves around the distribution of a dataset to return a probability for a given example.

### https://machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks/

Loss Functions is divided into three parts; they are:

1. Regression Loss Functions
   1. Mean Squared Error Loss
   2. Mean Squared Logarithmic Error Loss
   3. Mean Absolute Error Loss
2. Binary Classification Loss Functions
   1. Binary Cross-Entropy
   2. Hinge Loss
   3. Squared Hinge Loss
3. Multi-Class Classification Loss Functions
   1. Multi-Class Cross-Entropy Loss
   2. Sparse Multiclass Cross-Entropy Loss
   3. Kullback Leibler Divergence Loss