 Linear Algebra Revision questions.

**Q1**. Define Point/Vector (2-D, 3-D, n-D)?  
**A1**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2830/introduction-to-vectors2-d-3-d-n-d-row-vector-and-column-vector/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q2**. How to calculate Dot product and angle between 2 vectors?  
**A2**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2831/dot-product-and-angle-between-2-vectors/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q3**. Define Projection, unit vector?  
**A3**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2832/projection-and-unit-vector/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q4**. Equation of a line (2-D), plane(3-D) and hyperplane (n-D)?  
**A4**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2833/equation-of-a-line-2-d-plane3-d-and-hyperplane-n-d-plane-passing-through-origin-normal-to-a-plane/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q5**. Distance of a point from a plane/hyperplane, half-spaces?  
**A5**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2834/distance-of-a-point-from-a-planehyperplane-half-spaces/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q6**. Equation of a circle (2-D), sphere (3-D) and hypersphere (n-D)?  
**A6**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2835/equation-of-a-circle-2-d-sphere-3-d-and-hypersphere-n-d/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q7**. Equation of an ellipse (2-D), ellipsoid (3-D) and hyperellipsoid (n-D)?  
**A7**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2836/equation-of-an-ellipse-2-d-ellipsoid-3-d-and-hyperellipsoid-n-d/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q8**. Square, Rectangle, Hyper-cube and Hyper-cuboid?  
**A8**. <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2837/square-rectangle/2/module-2-data-science-exploratory-data-analysis-and-data-visualization>  
  
**Q9**. Check these videos 3b1b LA: <https://www.youtube.com/watch?v=fNk_zzaMoSs&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab>  
  
**Q10**. Check these videos 3b1b Calculus: <https://www.youtube.com/watch?v=WUvTyaaNkzM&list=PLZHQObOWTQDMsr9K-rj53DwVRMYO3t5Yr>  
  
**Q11**. Check these videos 3b1b calculus: <https://www.youtube.com/watch?v=p_di4Zn4wz4&list=PLZHQObOWTQDNPOjrT6KVlfJuKtYTftqH6>

Revision questions Probability and Statistics

1. hat is PDF?(https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2842/gaussiannormal-distribution-and-its-pdfprobability-density-function/2/module-2-data-science-exploratory-data-analysis-and-data-visualization)
3. What is CDF?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2843/cdfcumulative-distribution-function-of-gaussiannormal-distribution/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
5. explain about 1-std-dev, 2-std-dev, 3-std-dev range?
7. What is Symmetric distribution, Skewness and Kurtosis?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2844/symmetric-distribution-skewness-and-kurtosis/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
9. How to do Standard normal variate (z) and standardization?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2845/standard-normal-variate-z-and-standardization/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
11. What is Kernel density estimation?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2846/kernel-density-estimation/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
13. Importance of Sampling distribution & Central Limit theoremhttps://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2847/sampling-distribution-central-limit-theorem/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
15. Importance of Q-Q Plot: Is a given random variable Gaussian distributed?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2848/q-q-plothow-to-test-if-a-random-variable-is-normally-distributed-or-not/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
17. What is Uniform Distribution and random number generatorshttps://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2850/how-to-randomly-sample-data-points-uniform-distribution/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
19. What Discrete and Continuous Uniform distributions?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2849/discrete-and-continuous-uniform-distributions/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
20. How to randomly sample data points?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2850/how-to-randomly-sample-data-points-uniform-distribution/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
22. Explain about Bernoulli and Binomial distribution?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2851/bernoulli-and-binomial-distribution/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
24. What is Log-normal  and power law distribution?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2852/log-normal-distribution/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
26. What is Power-law & Pareto distributions: PDF, exampleshttps://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2853/power-law-distribution/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
28. Explain about Box-Cox/Power transform?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2854/box-cox-transform/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
30. What is Co-variance?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2855/co-variance/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
32. Importance of Pearson Correlation Coefficient?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2856/pearson-correlation-coefficient/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
34. Importance Spearman Rank Correlation Coefficient?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2857/spearman-rank-correlation-coefficient/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
36. Correlation vs Causation?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2858/correlation-vs-causation/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
38. What is Confidence Intervals?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2859/confidence-interval-ci-introduction/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
40. Confidence Interval vs Point estimate?
41. Explain about Hypothesis testing?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2863/hypothesis-testing-methodology-null-hypothesis-p-value/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
43. Define Hypothesis Testing methodology, Null-hypothesis, test-statistic, p-value?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2863/hypothesis-testing-methodology-null-hypothesis-p-value/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
45. How to do K-S Test for similarity of two distributions?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2865/k-s-test-for-similarity-of-two-distributions/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

Interview Questions on Probability and statistics

1. What is a random variable?
2. What are the conditions for a function to be a probability mass function?(http://www.statisticshowto.com/probability-mass-function-pmf/)
3. What are the conditions for a function to be a probability density function ?(Covered in our videos)
4. What is conditional probability?
5. State the Chain rule of conditional probabilities?(https://en.wikipedia.org/wiki/Chain\_rule\_(probability))
6. What are the conditions for independence and conditional independence of two random variables?(https://math.stackexchange.com/questions/22407/independence-and-conditional-independence-between-random-variables)
7. What are expectation, variance and covariance?(Covered in our videos)
8. Compare covariance and independence?(https://stats.stackexchange.com/questions/12842/covariance-and-independence)
9. What is the covariance for a vector of random variables?(https://math.stackexchange.com/questions/2697376/find-the-covariance-matrix-of-a-vector-of-random-variables)
10. What is a Bernoulli distribution?
11. What is a normal distribution?
12. What is the central limit theorem?
13. Write the formula for Bayes rule?
14. If two random variables are related in a deterministic way, how are the PDFs related?
15. What is Kullback-Leibler (KL) divergence?
16. Can KL divergence be used as a distance measure?
17. What is Bayes’ Theorem? How is it useful in a machine learning context?
18. Why is “Naive” Bayes naive?
19. What’s a Fourier transform?
20. What is the difference between covariance and correlation?
21. Is it possible capture the correlation between continuous and categorical variable? If yes, how?
22. What is the Box-Cox transformation used for?
23. What does P-value signify about the statistical data?
24. A test has a true positive rate of 100% and false positive rate of 5%. There is a population with a 1/1000 rate of having the condition the test identifies. Considering a positive test, what is the probability of having that condition?
25. How you can make data normal using Box-Cox transformation?
26. Explain about the box cox transformation in regression models.
27. What is the difference between skewed and uniform distribution?
28. What do you understand by Hypothesis in the content of Machine Learning?
29. How will you find the correlation between a categorical variable and a continuous variable ?
30. How to sample from a Normal Distribution with known mean and variance?

(t-SNE)T-distributed Stochastic Neighbourhood Embedding

1. What is dimensionality reduction? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2878/what-is-dimensionality-reduction/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
3. Explain Principal Component Analysis?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2889/geometric-intuition-of-pca/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
5. Importance of PCA?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2888/why-learn-pca/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
7. Limitations of PCA?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2894/limitations-of-pca/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
9. What is t-SNE?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2898/what-is-t-sne/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
11. What is Crowding problem?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2901/crowding-problem/2/module-2-data-science-exploratory-data-analysis-and-data-visualization
13. How to apply t-SNE and interpret its output?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2902/how-to-apply-t-sne-and-interpret-its-output/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

Interview Questions on Dimensionality Reduction

   
You are given a train data set having 1000 columns and 1 million rows. The data set is based on a classification problem. Your manager has asked you to reduce the dimension of this data so that model computation time can be reduced. Your machine has memory constraints. What would you do? (You are free to make practical assumptions.)(https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/)

 Is rotation necessary in PCA? If yes, Why? https://google-interview-hacks.blogspot.com/2017/04/is-rotation-necessary-in-pca-if-yes-why.html

 You are given a data set. The data set contains many variables, some of which are highly correlated and you know about it. Your manager has asked you to run PCA. Would you remove correlated variables first? Why?(<https://www.linkedin.com/pulse/questions-machine-learning-statistics-can-you-answer-saraswat/>)

**Revision Questions of KNN**

1. Explain about K-Nearest Neighbors?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2927/k-nearest-neighbours-geometric-intuition-with-a-toy-example/3/module-3-foundations-of-natural-language-processing-and-machine-learning
3. Failure cases of KNN?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2928/failure-cases-of-knn/3/module-3-foundations-of-natural-language-processing-and-machine-learning
5. Define Distance measures: Euclidean(L2) , Manhattan(L1), Minkowski,  Hamminghttps://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2929/distance-measures-euclideanl2-manhattanl1-minkowski-hamming/3/module-3-foundations-of-natural-language-processing-and-machine-learning
7. What is Cosine Distance & Cosine Similarity?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2930/cosine-distance-cosine-similarity/3/module-3-foundations-of-natural-language-processing-and-machine-learning
9. How to measure the effectiveness of k-NN?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2931/how-to-measure-the-effectiveness-of-k-nn/3/module-3-foundations-of-natural-language-processing-and-machine-learning
10. Limitations of KNN?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2933/knn-limitations/3/module-3-foundations-of-natural-language-processing-and-machine-learning
12. How to handle Overfitting and Underfitting in KNN?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2935/overfitting-and-underfitting/3/module-3-foundations-of-natural-language-processing-and-machine-learning
14. Need for Cross validation?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2936/need-for-cross-validation/3/module-3-foundations-of-natural-language-processing-and-machine-learning
16. What is K-fold cross validation?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2937/k-fold-cross-validation/3/module-3-foundations-of-natural-language-processing-and-machine-learning
18. What is Time based splitting?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2940/time-based-splitting/3/module-3-foundations-of-natural-language-processing-and-machine-learning
20. Explain k-NN for regression?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2941/k-nn-for-regression/3/module-3-foundations-of-natural-language-processing-and-machine-learning
22. Weighted k-NN ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2942/weighted-k-nn/3/module-3-foundations-of-natural-language-processing-and-machine-learning
24. How to build a kd-tree.?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2945/how-to-build-a-kd-tree/3/module-3-foundations-of-natural-language-processing-and-machine-learning
26. Find nearest neighbors using kd-treehttps://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2946/find-nearest-neighbours-using-kd-tree/3/module-3-foundations-of-natural-language-processing-and-machine-learning
28. What is Locality sensitive Hashing (LSH)?(
29. Hashing vs LSH?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2949/hashing-vs-lsh/3/module-3-foundations-of-natural-language-processing-and-machine-learning
31. LSH for cosine similarity?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2950/lsh-for-cosine-similarity/3/module-3-foundations-of-natural-language-processing-and-machine-learning
33. LSH for euclidean distance?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2951/lsh-for-euclidean-distance/3/module-3-foundations-of-natural-language-processing-and-machine-learning

Interview Questions on K-NN(K Nearest Neighbour)

1. In k-means or kNN, we use euclidean distance to calculate the distance between nearest neighbours. Why not manhattan distance ?(https://www.analyticsvidhya.com/blog/2017/09/30-questions-test-k-nearest-neighbors-algorithm/)
2. How to test and know whether or not we have overfitting problem?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/how-to-determine-overfitting-and-underfitting/)
3. How is kNN different from k-means clustering?(https://stats.stackexchange.com/questions/56500/what-are-the-main-differences-between-k-means-and-k-nearest-neighbours)
4. Can you explain the difference between a Test Set and a Validation Set?(https://stackoverflow.com/questions/2976452/whats-is-the-difference-between-train-validation-and-test-set-in-neural-netwo)
5. How can you avoid overfitting in KNN?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/how-to-determine-overfitting-and-underfitting/>)

 Classification algorithms in various situations Revision questions.

1. What is Imbalanced and  balanced dataset.  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2958/imbalanced-vs-balanced-dataset/3/module-3-foundations-of-natural-language-processing-and-machine-learning
2. Define Multi-class classification?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2959/multi-class-classification/3/module-3-foundations-of-natural-language-processing-and-machine-learning
3. Explain Impact of Outliers?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2962/impact-of-outliers/3/module-3-foundations-of-natural-language-processing-and-machine-learning
4. What is Local Outlier Factor?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2963/local-outlier-factor-simple-solution-mean-distance-to-knn/3/module-3-foundations-of-natural-language-processing-and-machine-learning
5. What is k-distance (A), N(A)  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2964/k-distance/3/module-3-foundations-of-natural-language-processing-and-machine-learning
6. Define reachability-distance(A, B)?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2965/reachability-distanceab/3/module-3-foundations-of-natural-language-processing-and-machine-learning
7. What is Local-reachability-density(A)?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2966/local-reachability-densitya/3/module-3-foundations-of-natural-language-processing-and-machine-learning
8. Define LOF(A)?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2967/local-outlier-factora/3/module-3-foundations-of-natural-language-processing-and-machine-learning
9. Impact of Scale & Column standardization?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2968/impact-of-scale-column-standardization/3/module-3-foundations-of-natural-language-processing-and-machine-learning
10. What is Interpretability?  
    https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2969/interpretability/3/module-3-foundations-of-natural-language-processing-and-machine-learning
11. Handling categorical and numerical features?  
    https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2971/handling-categorical-and-numerical-features/3/module-3-foundations-of-natural-language-processing-and-machine-learning
12. Handling missing values by imputation?  
    https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2977/handling-missing-values-by-imputation/3/module-3-foundations-of-natural-language-processing-and-machine-learning
13. Bias-Variance tradeoff?  
    https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2973/bias-variance-tradeoff/3/module-3-foundations-of-natural-language-processing-and-machine-learning

Interview Questions on Performance Measurement Models

1. Which is more important to you– model accuracy, or model performance?
2. Can you cite some examples where a false positive is important than a false negative?
3. Can you cite some examples where a false negative important than a false positive?
4. Can you cite some examples where both false positive and false negatives are equally important?
5. What is the most frequent metric to assess model accuracy for classification problems?
6. Why is Area Under ROC Curve (AUROC) better than raw accuracy as an out-of- sample evaluation metric?

Naïve-Bayes **Revision Questions :**

1. Bayes Theorem problem: https://youtu.be/LadMzl8MaXM
2. More Bayes Theorem problems: https://www.math.upenn.edu/~mmerling/math107%20docs/practice%20on%20Bayes%20solutions.pdf http://gtribello.github.io/mathNET/bayes-theorem-problems.html http://wwwf.imperial.ac.uk/~ayoung/m2s1/WorkedExamples1.pdf
3. What is Conditional probability? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2989/conditional-probability/3/module-3-foundations-of-natural-language-processing-and-machine-learnin
4. Define Independent vs Mutually exclusive events?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2990/independent-vs-mutually-exclusive-events/3/module-3-foundations-of-natural-language-processing-and-machine-learning
5. Explain Bayes Theorem with example? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2991/bayes-theorem-with-examples/3/module-3-foundations-of-natural-language-processing-and-machine-learning
6. How to apply Naive Bayes on Text data? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2995/naive-bayes-on-text-data/3/module-3-foundations-of-natural-language-processing-and-machine-learning
7. What is Laplace/Additive Smoothing? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2996/laplaceadditive-smoothing/3/module-3-foundations-of-natural-language-processing-and-machine-learning
8. Explain Log-probabilities for numerical stability? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2997/log-probabilities-for-numerical-stability/3/module-3-foundations-of-natural-language-processing-and-machine-learning
9. In Naive bayes how to handle  Bias and Variance tradeoff? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2998/bias-and-variance-tradeoff/3/module-3-foundations-of-natural-language-processing-and-machine-learning
10. What Imbalanced data? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3000/imbalanced-data/3/module-3-foundations-of-natural-language-processing-and-machine-learning
11. What is Outliers and how to handle outliers? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3001/outliers/3/module-3-foundations-of-natural-language-processing-and-machine-learning
12. How to handle Missing values? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3002/missing-values/3/module-3-foundations-of-natural-language-processing-and-machine-learning
13. How to Handling Numerical features (Gaussian NB) https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3003/handling-numerical-features-gaussian-nb/3/module-3-foundations-of-natural-language-processing-and-machine-learning
14. Define Multiclass classification.? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3004/multiclass-classification/3/module-3-foundations-of-natural-language-processing-and-machine-learning

 Solving Optimization Problems **Revison  Questions**

1. Explain about Logistic regression?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3011/geometric-intuition-of-logistic-regression/3/module-3-foundations-of-natural-language-processing-and-machine-learning
3. What is Sigmoid function & Squashing ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3012/sigmoid-function-squashing/3/module-3-foundations-of-natural-language-processing-and-machine-learning
5. Explain about Optimization problem in logistic regression. https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3013/mathematical-formulation-of-objective-function/3/module-3-foundations-of-natural-language-processing-and-machine-learning
7. Expalain Importance of Weight vector in logistic regression.https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3014/weight-vector/3/module-3-foundations-of-natural-language-processing-and-machine-learning
9. L2 Regularization: Overfitting and Underfittinghttps://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3015/l2-regularization-overfitting-and-underfitting/3/module-3-foundations-of-natural-language-processing-and-machine-learning
11. L1 regularization and sparsity. https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3016/l1-regularization-and-sparsity/3/module-3-foundations-of-natural-language-processing-and-machine-learning
13. What is Probabilistic Interpretation: Gaussian Naive Bayes ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3017/probabilistic-interpretation-gaussian-naive-bayes/3/module-3-foundations-of-natural-language-processing-and-machine-learning
14. Explain about Hyperparameter search: Grid Search and Random Search ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3019/hyperparameters-and-random-search/3/module-3-foundations-of-natural-language-processing-and-machine-learning
16. What is Column Standardization.?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3020/column-standardization/3/module-3-foundations-of-natural-language-processing-and-machine-learning
18. Explain about Collinearity of features?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3022/collinearity-of-features/3/module-3-foundations-of-natural-language-processing-and-machine-learning
20. Find Train & Run time space and time complexity of Logistic regression?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3023/testrun-time-space-and-time-complexity/3/module-3-foundations-of-natural-language-processing-and-machine-learning

 Interview Questions on Logistic Regression and Linear Regression

1. Outliers and Loss Functions: https://youtu.be/jiOBCCZCtug  
2. After analyzing the model, your manager has informed us that your regression model is suffering from multicollinearity. How would you check if he’s true? Without losing any information, can you still build a better model?(https://google-interview-hacks.blogspot.in/2017/04/after-analyzing-model-your-manager-has.html)  
3. What are the basic assumptions to be made for linear regression?(https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/geometric-intuition-1-2-copy-8/)  
4. What is the difference between stochastic gradient descent (SGD) and gradient descent (GD)?(https://stats.stackexchange.com/questions/317675/gradient-descent-gd-vs-stochastic-gradient-descent-sgd)

5. When would you use GD over SDG, and vice-versa?(https://elitedatascience.com/machine-learning-interview-questions-answers)

6. How do you decide whether your linear regression model fits the data?(https://www.researchgate.net/post/What\_statistical\_test\_is\_required\_to\_assess\_goodness\_of\_fit\_of\_a\_linear\_or\_nonlinear\_regression\_equation)

7. Is it possible to perform logistic regression with Microsoft Excel?(https://www.youtube.com/watch?v=EKRjDurXau0)

8. When will you use classification over regression?(https://www.quora.com/When-will-you-use-classification-over-regression)

9. Why isn't Logistic Regression called Logistic Classification?(Refer :https://stats.stackexchange.com/questions/127042/why-isnt-logistic-regression-called-logistic-classification/127044)  
**More  External Resources:**  
1.<https://www.analyticsvidhya.com/blog/2017/08/skilltest-logistic-regression/>  
2.<https://www.listendata.com/2017/03/predictive-modeling-interview-questions.html>  
3.<https://www.analyticsvidhya.com/blog/2017/07/30-questions-to-test-a-data-scientist-on-linear-regression/> 4.<https://www.analyticsvidhya.com/blog/2016/12/45-questions-to-test-a-data-scientist-on-regression-skill-test-regression-solution/>  
[5. https://www.listendata.com/2018/03/regression-analysis.html](https://www.listendata.com/2018/03/regression-analysis.html)

 Support Vector Machines (SVM) **Revision Questions:**

1. Explain About SVM?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3047/geometric-intution/4/module-4-machine-learning-ii-supervised-learning-models
3. What is Hinge Loss?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3050/loss-function-hinge-loss-based-interpretation/4/module-4-machine-learning-ii-supervised-learning-models
5. Dual form of SVM formulation.?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3051/dual-form-of-svm-formulation/4/module-4-machine-learning-ii-supervised-learning-models
7. What is Kernel trick.?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3052/kernel-trick/4/module-4-machine-learning-ii-supervised-learning-models
9. What is Polynomial kernel.?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3053/polynomial-kernel/4/module-4-machine-learning-ii-supervised-learning-models
10. What is RBF-Kernel.?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3054/rbf-kernel/4/module-4-machine-learning-ii-supervised-learning-models
12. Explain about Domain specific Kernels. ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3055/domain-specific-kernels/4/module-4-machine-learning-ii-supervised-learning-models
14. Find Train and run time complexities for SVM?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3056/train-and-run-time-complexities/4/module-4-machine-learning-ii-supervised-learning-models
16. Explain about SVM Regression. ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3058/svm-regression/4/module-4-machine-learning-ii-supervised-learning-models

Support Vector Machines (SVM) -Interview questions.

1. Give some situations where you will use an SVM over a RandomForest Machine Learning algorithm and vice-versa.(https://datascience.stackexchange.com/questions/6838/when-to-use-random-forest-over-svm-and-vice-versa)
2. What is convex hull ?(https://en.wikipedia.org/wiki/Convex\_hull)
3. What is a large margin classifier?
4. Why SVM is an example of a large margin classifier?
5. SVM being a large margin classifier, is it influenced by outliers? (Yes, if C is large, otherwise not)
6. What is the role of C in SVM?
7. In SVM, what is the angle between the decision boundary and theta?
8. What is the mathematical intuition of a large margin classifier?
9. What is a kernel in SVM? Why do we use kernels in SVM?
10. What is a similarity function in SVM? Why it is named so?
11. How are the landmarks initially chosen in an SVM? How many and where?
12. Can we apply the kernel trick to logistic regression? Why is it not used in practice then?
13. What is the difference between logistic regression and SVM without a kernel? (Only in implementation – one is much more efficient and has good optimization packages)
14. How does the SVM parameter C affect the bias/variance trade off? (Remember C = 1/lambda; lambda increases means variance decreases)
15. How does the SVM kernel parameter sigma^2 affect the bias/variance trade off?
16. Can any similarity function be used for SVM? (No, have to satisfy Mercer’s theorem)
17. Logistic regression vs. SVMs: When to use which one? ( Let's say n and m are the number of features and training samples respectively. If n is large relative to m use log. Reg. or SVM with linear kernel, If n is small and m is intermediate, SVM with Gaussian kernel, If n is small and m is massive, Create or add more features then use log. Reg. or SVM without a kernel)
18. What is the difference between supervised and unsupervised machine learning?

**External Resources:** 1.<https://www.analyticsvidhya.com/blog/2017/10/svm-skilltest/>

 Decision Trees revision questions.

1. How to Building a decision Tree?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3064/geometric-intuition-of-decision-tree-axis-parallel-hyperplanes/4/module-4-machine-learning-ii-supervised-learning-models
3. What is Entropy?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3066/building-a-decision-treeentropy/4/module-4-machine-learning-ii-supervised-learning-models
5. What is information Gain ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3067/building-a-decision-treeinformation-gain/4/module-4-machine-learning-ii-supervised-learning-models
7. What is Gini Impurity?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3068/building-a-decision-tree-gini-impurity/4/module-4-machine-learning-ii-supervised-learning-models
9. How to Constructing a DT. ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3069/building-a-decision-tree-constructing-a-dt/4/module-4-machine-learning-ii-supervised-learning-models
11. Importance of Splitting numerical features.?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3070/building-a-decision-tree-splitting-numerical-features/4/module-4-machine-learning-ii-supervised-learning-models
13. How to handle Overfitting and Underfitting in DT?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3073/overfitting-and-underfitting/4/module-4-machine-learning-ii-supervised-learning-models
15. What are Train and Run time complexity for DT?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3074/train-and-run-time-complexity/4/module-4-machine-learning-ii-supervised-learning-models
17. How to implement Regression using Decision Trees?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3075/regression-using-decision-trees/4/module-4-machine-learning-ii-supervised-learning-models

 Interview Questions on decision Trees

**Self Learning:**

1. You are working on a time series data set. You manager has asked you to build a high accuracy model. You start with the decision tree algorithm, since you know it works fairly well on all kinds of data. Later, you tried a time series regression model and got higher accuracy than decision tree model. Can this happen? Why?(Refer :https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/)
2. Running a binary classification tree algorithm is the easy part. Do you know how does a tree splitting takes place i.e. how does the tree decide which variable to split at the root node and succeeding nodes?(Refer:https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/)

  External Resources: 1.[https://vitalflux.com/decision-tree-algorithm-concepts-interview-question](https://vitalflux.com/decision-tree-algorithm-concepts-interview-questions-set-1/)

Ensemble Models **Revision Questions:**

1. What are ensembles?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3081/what-are-ensembles/4/module-4-machine-learning-ii-supervised-learning-models
3. What is Bootstrapped Aggregation (Bagging) ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3082/bootstrapped-aggregation-bagging-intuition/4/module-4-machine-learning-ii-supervised-learning-models
5. Explain about Random Forest and their construction?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3083/random-forest-and-their-construction/4/module-4-machine-learning-ii-supervised-learning-models
7. Explain about Boosting?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3089/boosting-intuition/4/module-4-machine-learning-ii-supervised-learning-models
9. What are Residuals, Loss functions and gradients ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3090/residuals-loss-functions-and-gradients/4/module-4-machine-learning-ii-supervised-learning-models
11. Explain about Gradient Boosting?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3091/gradient-boosting/4/module-4-machine-learning-ii-supervised-learning-models
13. What is Regularization by Shrinkage?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3092/regularization-by-shrinkage/4/module-4-machine-learning-ii-supervised-learning-models
15. Explain about XGBoost?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3094/xgboost-boosting-randomization/4/module-4-machine-learning-ii-supervised-learning-models
17. Explain about AdaBoost?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3095/adaboost-geometric-intuition/4/module-4-machine-learning-ii-supervised-learning-models
19. How do you implement Stacking models?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3096/stacking-models/4/module-4-machine-learning-ii-supervised-learning-models
21. Explain about cascading classifiers. ?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3097/cascading-classifiers/4/module-4-machine-learning-ii-supervised-learning-models

DBSCAN (Density based clustering) Technique **Revision Questions:**

1. What is K-means? How can you select K for K-means?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3263/k-means-algorithm/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
3. How is KNN different from k-means clustering?
4. Explain about Hierarchical clustering?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3271/agglomerative-divisive-dendrograms/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
6. Limitations of Hierarchical clustering?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3275/limitations-of-hierarchical-clustering/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
8. Time complexity of Hierarchical clustering?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3274/time-and-space-complexity/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
10. Explain about DBSCAN?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3282/dbscan-algorithm/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
12. Advantages and Limitations of DBSCAN?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3284/advantages-and-limitations-of-dbscan/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies

 Recommender Systems and Matrix Factorization : **Revision Questions:**

1. Explain about Content based and Collaborative Filtering?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3290/content-based-vs-collaborative-filtering/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
2. What is PCA, SVD?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3292/matrix-factorization-pca-svd/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
3. What is NMF?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3293/matrix-factorization-nmf/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
4. How to do MF for Collaborative filtering ?  
   <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online->
5. course/3294/matrix-factorization-for-collaborative-filtering/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
6. How to do MF for feature engineering.?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3295/matrix-factorization-for-feature-engineering/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
7. Explain relation between Clustering And  MF?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3296/clustering-as-mf/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
8. What is Hyperparameter tuning. ?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3297/hyperparameter-tuning/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
9. Explain about Cold Start problem.?  
   https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3299/cold-start-problem/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
10. How to solve Word Vectors using MF?  
    https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3300/word-vectors-as-mf/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies
11. Explain about Eigenfaces. ?  
    <https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/3301/eigen-faces/6/module-7-data-miningunsupervised-learning-and-recommender-systems-real-world-case-studies>

 Interview Questions on Recommender Systems and Matrix Factorization.

**Self Learning:**

1. How would you implement a recommendation system for our company’s users?(https://www.infoworld.com/article/3241852/machine-learning/how-to-implement-a-recommender-system.html)
2. How would you approach the “Netflix Prize” competition?(Refer http://blog.echen.me/2011/10/24/winning-the-netflix-prize-a-summary/)
3. ‘People who bought this, also bought…’ recommendations seen on amazon is a result of which algorithm?(Please refer Apparel recommendation system case study,Refer:https://measuringu.com/affinity-analysis/)

 Deep Learning:Neural Networks.

Google Interview Question + Solution : <https://youtu.be/vwYmrKDvnlg>

Deep Learning: Convolutional Neural Nets.

<https://soundcloud.com/applied-ai-course/how-to-build-a-face>

 Interview Questions on Deep Learning

## Basics of Natural Language Processing(NLP):

1. 1.Explain about Bag of Words?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/bag-of-words-bow/>)
2. Explain about Text Preprocessing: Stemming, Stop-word removal, Tokenization, Lemmatization.(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/text-preprocessing-stemming-stop-word-removal-tokenization-lemmatization/>)
3. Explain about uni-gram, bi-gram, n-grams.?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/uni-gram-bi-gram-n-grams/>)
4. What is tf-idf (term frequency- inverse document frequency)(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/tf-idf-term-frequency-inverse-document-frequency/>)
5. Why use log in IDF?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/why-use-log-in-idf/>)
6. Explain about Word2Vec.?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/word2vec/>)
7. Explain about Avg-Word2Vec, tf-idf weighted Word2Vec?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/avg-word2vec-tf-idf-weighted-word2vec/>)
8. Explain about Multi-Layered Perceptron (MLP)?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/multi-layered-perceptron-mlp/>)
9. How to train a single-neuron model?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/training-a-single-neuron-model/>)
10. How to Train an MLP using Chain rule ?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/training-an-mlp-2/>)
11. How to Train an MLP using Memoization?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/training-an-mlp/>)
13. Explain about Backpropagation algorithm?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/backpropagation/>)
14. Describe about Vanishing and Exploding Gradient problem?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/vanishing-gradient-problem-2/>)
15. Explain about Bias-Variance tradeoff in neural Networks?(<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/bias-variance-tradeoff-23/>)

**Deep Learning:**

1. What is sampled softmax?
2. Why is it difficult to train a RNN with SGD?
3. How do you tackle the problem of exploding gradients? (By gradient clipping)
4. What is the problem of vanishing gradients? (RNN doesn't tend to remember much things from the past)
5. How do you tackle the problem of vanishing gradients? (By using LSTM)
6. Explain the memory cell of a LSTM. (LSTM allows forgetting of data and using long memory when appropriate.)
7. What type of regularization do one use in LSTM?
8. What is the problem with sigmoid during backpropagation? (Very small, between 0.25 and zero.)
9. What is transfer learning?
10. What is backpropagation through time? (BPTT)
11. What is the difference between LSTM and GRU?
12. Explain Gradient Clipping.
13. Adam and RMSProp adjust the size of gradients based on previously seen gradients. Do they inherently perform gradient clipping? If no, why?

**External sources** <https://www.analyticsvidhya.com/blog/2017/01/must-know-questions-deep-learning/>

# External resources for Interview Questions

**References:**   <https://medium.com/acing-ai/salesforce-ai-interview-questions-acing-the-ai-interview-75e177c4734> <https://medium.com/acing-ai/microsoft-ai-interview-questions-acing-the-ai-interview-be6972f790ea> <https://medium.com/acing-ai/apple-ai-interview-questions-acing-the-ai-interview-803a65b0e795> <https://medium.com/acing-ai/amazon-ai-interview-questions-acing-the-ai-interview-3ed4e671920f> <https://medium.com/acing-ai/uber-ai-interview-questions-acing-the-ai-interview-9532794bc057> <https://medium.com/acing-ai/steps-to-ace-the-ai-interview-part-1-298249080e59> <https://medium.com/acing-ai/steps-to-ace-the-ai-interview-part-2-b25f91582f5f> <https://medium.com/acing-ai/google-ai-interview-questions-acing-the-ai-interview-1791ad7dc3ae> <https://medium.com/acing-ai/facebook-ai-interview-questions-acing-the-ai-interview-5982add0af55> <https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/> <https://www.analyticsvidhya.com/blog/2017/09/30-questions-test-k-nearest-neighbors-algorithm/> <https://www.analyticsvidhya.com/blog/2017/08/skilltest-logistic-regression/> <https://www.listendata.com/2017/03/predictive-modeling-interview-questions.html> <https://www.analyticsvidhya.com/blog/2017/07/30-questions-to-test-a-data-scientist-on-linear-regression/> <https://www.analyticsvidhya.com/blog/2016/12/45-questions-to-test-a-data-scientist-on-regression-skill-test-regression-solution/> https://medium.com/acing-ai/adobe-ai-interview-questions-acing-the-ai-interview-ef7a8099110b <https://www.listendata.com/2018/03/regression-analysis.html> <https://www.analyticsvidhya.com/blog/2017/10/svm-skilltest/> [https://vitalflux.com/decision-tree-algorithm-concepts-interview-question](https://vitalflux.com/decision-tree-algorithm-concepts-interview-questions-set-1/) <https://www.analyticsvidhya.com/blog/2017/09/30-questions-test-tree-based-models/> <https://www.analyticsvidhya.com/blog/2017/01/must-know-questions-deep-learning/> <https://www.kdnuggets.com/tag/interview-questions>