Interview Questions

# Yet another questions list

\* Why do we need the 8209th questions list on github? The questions here sorted by different area and focus on testing your actual knowledge instead of checking your memorization.

\* How do you pick questions? A good question can distinguish candidate that \*\*understand\*\* ve candidates that \*\*knows\*\* the concept. There are two level of questions: fundamental and advance, depend on your timeline you might want to priortize your learning accordingly.

| Question | Level |

| ------------- | ------------- |

| Underlying decision tree algorithm, what's the difference between regression and classification? | Fundamental |

| What is the difference between logistic regression vs multilayer perceptron? | Advance |

| [Why L1 sparse?](https://stats.stackexchange.com/questions/45643/why-l1-norm-for-sparse-models) | Advance |

| Why do RNNs have a tendency to suffer from exploding/vanishing gradient? | Fundamental |

| What is the problem with sigmoid during backprop? | Advance |

| What is the problem with Relu? Why do we need Leaky Relu? | Fundamental |

| What are the assumptions of linear/logistic regression? | Fundamental |

| If some features have strong colinearity does it make model performance worse? | Fundamental |

| What does the decision boundary of logistic regression look like? | Fundamental |

# List 99 ML questions

## Basic stats questions

1. What is the normal distribution and its functions?

2. What is the variance?

3. What is stddev?

4. What is exponential distribution?

5. What is the binomial distribution?

6. What is Bernoulli distribution?

7. What is the multinomial distribution?

8. What is the lognormal distribution?

9. What is logistic distribution?

10. P value and test hypothesis.

11. Bias coin, unbiased coin posterior probability. Observe the head, what is the likelihood of

coin bias.

12. What is hypothesis testing? chi-squared test: determine if any statistically significant

difference between expected frequencies and observed frequencies in one or more

categories. More popularly used in medicine.

13. What is the gamma distribution and its functions?

14. Poisson distribution and its function.

15. What is the Central Limit Theorem and why is it important?

16. What is sampling? How many sampling methods do you know?

17. Basic Bayes Theorem. P(A|B) = P(B|A) \* P(A) / P(B). Conditional prob and marginal prob

18. What is an example of a data set with a non-Gaussian distribution?

19. Given a generator of unbiased Bernoulli numbers (0 or 1 with p=0.5), create a biased Bernoulli trial generator (generate 0 or 1 with the specified 0 < p < 1)

## Machine Learning fundamentals

20. Explain collinearity: how is it possible to have negative coefficient in glm?

21. Assumptions about linear regression: 3 residual errors follow a normal distribution and

independent with each other, output and input have linear distribution, low collinearity

between input features.

22. Explain the concept of bias vs variance

23. Assumptions about logistic regression?

24. Overfitting how to avoid

25. Type of regularization, which one easier to use: L1 and L2. L1 is easier.

26. What are the different metrics to classify the dataset?

27. explain bagging vs boosting

28. Is random forest bagging or boosting?.

29. how to handle unbalanced data.

30. how do you inspect missing data and when are they important

31. how to compare two regressions.

32. What is k-mean? What is kmean loss function? what kind of distance metric would you choose, what if different features have a different dynamic range Kmeans is an algorithm to group similar data points into the same groups. The loss function used L2 norm: sum

of the square of the distance to the centroid. Similar to the Expectation-Maximization algorithm.

33. general ML questions like generative v.s.

34. What is the difference between type I vs type II error?

35. What is the linear regression? What do the terms p-value, coefficient, and r-squared

value mean? What is the significance of each of these components?

36. What is selection bias? It is the bias introduced by the selection from groups that is not

achieved by randomness.

37. What is regularization? How does it solve bias, variance problems?

38. What is the learning curve tool?

39. What is the meaning of the contour visualization of the cost function in linear regression?

40. Explain logistic regression?

a. Why is the name logistic function?

b. Interpretation of logistic function

i. Why do we need continuous output of the logistics function?

c. What does the decision boundary look like?

d. What kind of normalization is needed?

e. Relationship with cross-entropy

f. What are the other logit families?

41. How does the optimization algorithm work i.e to minimize the cost function?

a. Gradient descent, Conjugate gradient, BFGS, L-BFGS

42. Explain SVM

a. What is the decision boundary?

b. What is a kernel trick?

c. What is the cost function?

d. How to tune hyperparameters to avoid underfit, overfit?

i. kernel choice

ii. c parameter

iii. Math formula

43. Explain the decision tree

a. Why do we need a decision tree?

b. How does it work? How do we split? How do we select which feature to split?

c. What is the meaning of the parameters?

d. What is the cost function? Why do we need that cost function? And in which case

do we need which

e. What does tree pruning mean?

44. explain random forest

a. How does it work?

b. How to tune parameters to avoid underfit/overfit?

c. Is it an ensemble of bagging?

45. Explain boosting

a. Basic intuition

b. When to do it?

c. What problem does it solve?

d. What are the parameters?

e. How to tune?

46. Explain PCA. What is physical intuition?

47. Explain MLP: kind of fully connected feedforward neural network that use sigmoid or tanh as activation functions, widely used for classification task similar to logistic regression.

48. Explain MLE

49. Explain MAP

50. what is cross-validation: the purpose is to estimate how a model performs on unseen data by partition data set into many sets and pick a model with higher

accuracy on the validation set.

51. SVM: how to choose the model and how to determine if a model is better than another

52. Explain how you select the best model?

53. Explain TPE hyperparameter optimization. explain hyper optimization: BayesOptimization

54. pros and cons of random forest and why. Pros: effective, better than tree, effective feature importance. Cons: not easy to interpret, prone to out of range input issue, less

effective with big data or nonsmooth assumption.

55. How GMM works (EM algorithm)

56. When using the Gaussian mixture model, how do you know it is applicable?

57. Describe some criteria for model selection? Why is dimension reduction important?

58. How do you choose the best model among all possible models? explain neural network assumptions of linear regression models if you have a large number of predictors how do you handle them?

59. What is the bias-variance tradeoff? How is XGBoost handling bias-variance tradeoff?

60. How do you find an anomaly in a distribution? How do you investigate that a certain trend in distribution is due to anomaly?

## Deep learning fundamentals

61. Explain all standard activation functions? tanh, relu, leakyRelu, sigmoid, softmax, softplus

62. How does a neural network with one layer and one input and output compare to logistic regression? It’s the same with logistic regression if the NN is using sigmoid function and logistic regression.

63. Explain RNN Problem. How does LSTM help solve it?

## Companies Interview questions:

64. Netflix question – When you split a population for A/B testing, what are some reasons you could see a significant difference in the control and variant groups?

65. Google question – Find the width of the confidence interval

66. Facebook question – If you draw 2 cards from a shuffled 52 card deck, what is the probability that you’ll have a pair?

67. Microsoft question – Generate 7 integers with equal probability from a function which returns 1/0 with probability p and (1-p).

68. Given an unfair coin with the probability of heads not equal to .5, what algorithm could

you use to create a list of random 1s and 0s?

69. Groupon question – You are on a number line and you can jump to one of the neighboring points with equal probability, with the exception of n=0 where you can’t go to negative numbers but have to come back to n=1. If you start at n=44, what is the expected number of steps to reach n=4444?

70. CA Technologies question – How do you design an algorithm for fraud detection?

71. LinkedIn question – Come up with some of the factors that could be used to produce certain algorithms (‘people you may know,’ and an algorithm to discover when a person is starting to search for new job). PYMK: uses many features to calculate a connection probability among two people. Some examples are: companies position overlap (How

many months/years both individuals work together at same company, similar/related

field), school education overlap, strong unconnected people in LinkedIn connection graph. (both individuals have many common friends) etc.

72. An important metric goes down, how would you dig into the causes?

73. Amazon question – Estimate the cumulative sum of the top 10 most profitable products

of the last 6 months for customers in Seattle.

74. Booking question – How can we automatically propose ‘good value deals’ to customers,

including hotels that don’t have a rating yet?

75. If you have a customer and want to decide whether they will “buy today” or “not buy

today” and you know 1. where they live, 2. their income, 3. their gender, 4. their

profession, how would you define a machine-learning algorithm to figure this out?

76. LinkedIn question – How would you design an A/B test for the homepage?

77. Netflix question – Given a month’s worth of login data from Netflix such as account\_id,

device\_id, and metadata concerning payments, how would you detect fraud? (identity

theft, payment fraud, etc.)

78. Booking question – How would you tag a listing as value for money? How would you

measure the “value”? What features would you select to explain the “value”?

79. Apple question – How do you take millions of users with 100s of transactions each,

amongst 10ks of products and group the users together in a meaningful segment?

80. Facebook question – How many high schools that people have listed on their profiles are

real? How do we find out, and deploy at scale, a way of finding invalid schools?

81. We have a product that is getting used differently by two different groups. What is your

hypothesis about why and how would you go about testing it?

82. Intuit question – How would you design a ranking system?

83. Uber question – Explain how network effects might influence your choice of how to

assign experimental/control units and measure your main outcome metrics

84. LinkedIn question – What product metrics do you construct? How do you tell if your experiment is successful?

85. What trends in the data indicate that a given market is healthy? What does price tell you?

86. Facebook question – Given a series of tables; write the SQL code you would need to count subpopulations through joins.

87. Pinterest question – Write a SQL query to count the number of unique users per day who logged in from both an iPhone and the web, where iPhone logs and weblogs are in

distinct relations.

88. Spotify question – Given a sample set of tables, write a SQL query to get a summary metric from those tables.

89. Twitter question – How can you illustrate a tree-based system with a SQL query?

90. If you have a table with a billion rows, how would you add a column inserting data from

the original source without affecting the user experience?

91. Facebook question – There is a table that tracks every time a user turns a feature on or off, with columns for user\_id, action (“on” or “off), date, and time. How many users turned the feature on today? How many users have ever turned the feature on? In a table that tracks the status of every user every day, how would you add today’s data to it?

92. Check if an integer is a palindrome (do not convert the integer to string)

93. Given 2 sorted arrays of integers, code to find a number from each array such that their

sum is closest to some integer K

94. Amazon question – Write a Python function that displays the first n Fibonacci numbers.

95. Cisco question – Merge 2 sorted linked list

96. eBay question – Given a function roll() that uniformly returns a double between 0 and 1 and an array/list of numbers of length N (no duplicates), create a function shuffle() that returns a permutation of equal probability.

97. How would you create/design/implement a certain algorithm from start to end?

98. LinkedIn question – Given a random generator that produces a number 1 to 5 uniformly,

write a function that produces a number from 1 to 7 uniformly

99. Generate a sorted vector from two sorted vectors.

# Other question list

\* [List of ML questions](https://github.com/Sroy20/machine-learning-interview-questions/blob/master/list\_of\_questions\_machine\_learning.md)

\* [List of DL questions](https://github.com/Sroy20/machine-learning-interview-questions/blob/master/list\_of\_questions\_deep\_learning.md)

\* [Other interview questions](https://github.com/Feynman687/Interviews/blob/master/StatML.md)

\* If you find this helpful, you can Sponsor this project. It's cool if you don't.

# Machine Learning - List of questions

## Learning Theory

1. Describe bias and variance with examples.

1. What is Empirical Risk Minimization?

1. What is Union bound and Hoeffding's inequality?

1. Write the formulae for training error and generalization error. Point out the differences.

1. State the uniform convergence theorem and derive it.

1. What is sample complexity bound of uniform convergence theorem?

1. What is error bound of uniform convergence theorem?

1. What is the bias-variance trade-off theorem?

1. From the bias-variance trade-off, can you derive the bound on training set size?

1. What is the VC dimension?

1. What does the training set size depend on for a finite and infinite hypothesis set? Compare and contrast.

1. What is the VC dimension for an n-dimensional linear classifier?

1. How is the VC dimension of a SVM bounded although it is projected to an infinite dimension?

1. Considering that Empirical Risk Minimization is a NP-hard problem, how does logistic regression and SVM loss work?

## Model and feature selection

1. Why are model selection methods needed?

1. How do you do a trade-off between bias and variance?

1. What are the different attributes that can be selected by model selection methods?

1. Why is cross-validation required?

1. Describe different cross-validation techniques.

1. What is hold-out cross validation? What are its advantages and disadvantages?

1. What is k-fold cross validation? What are its advantages and disadvantages?

1. What is leave-one-out cross validation? What are its advantages and disadvantages?

1. Why is feature selection required?

1. Describe some feature selection methods.

1. What is forward feature selection method? What are its advantages and disadvantages?

1. What is backward feature selection method? What are its advantages and disadvantages?

1. What is filter feature selection method and describe two of them?

1. What is mutual information and KL divergence?

1. Describe KL divergence intuitively.

## Curse of dimensionality

1. Describe the curse of dimensionality with examples.

1. What is local constancy or smoothness prior or regularization?

## Universal approximation of neural networks

1. State the universal approximation theorem? What is the technique used to prove that?

1. What is a Borel measurable function?

1. Given the universal approximation theorem, why can't a MLP still reach a arbitrarily small positive error?

## Deep Learning motivation

1. What is the mathematical motivation of Deep Learning as opposed to standard Machine Learning techniques?

1. In standard Machine Learning vs. Deep Learning, how is the order of number of samples related to the order of regions that can be recognized in the function space?

1. What are the reasons for choosing a deep model as opposed to shallow model? (1. Number of regions O(2^k) vs O(k) where k is the number of training examples 2. # linear regions carved out in the function space depends exponentially on the depth. )

1. How Deep Learning tackles the curse of dimensionality?

## Support Vector Machine

1. How can the SVM optimization function be derived from the logistic regression optimization function?

1. What is a large margin classifier?

1. Why SVM is an example of a large margin classifier?

1. SVM being a large margin classifier, is it influenced by outliers? (Yes, if C is large, otherwise not)

1. What is the role of C in SVM?

1. In SVM, what is the angle between the decision boundary and theta?

1. What is the mathematical intuition of a large margin classifier?

1. What is a kernel in SVM? Why do we use kernels in SVM?

1. What is a similarity function in SVM? Why it is named so?

1. How are the landmarks initially chosen in an SVM? How many and where?

1. Can we apply the kernel trick to logistic regression? Why is it not used in practice then?

1. 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)

1. How does the SVM parameter C affect the bias/variance trade off? (Remember C = 1/lambda; lambda increases means variance decreases)

1. How does the SVM kernel parameter sigma^2 affect the bias/variance trade off?

1. Can any similarity function be used for SVM? (No, have to satisfy Mercer’s theorem)

1. 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 fetaures then use log. Reg. or SVM without a kernel)

## Bayesian Machine Learning

1. What are the differences between “Bayesian” and “Freqentist” approach for Machine Learning?

1. Compare and contrast maximum likelihood and maximum a posteriori estimation.

1. How does Bayesian methods do automatic feature selection?

1. What do you mean by Bayesian regularization?

1. When will you use Bayesian methods instead of Frequentist methods? (Small dataset, large feature set)

## Regularization

1. What is L1 regularization?

1. What is L2 regularization?

1. Compare L1 and L2 regularization.

1. Why does L1 regularization result in sparse models? [here](https://stats.stackexchange.com/questions/45643/why-l1-norm-for-sparse-models)

## Evaluation of Machine Learning systems

1. What are accuracy, sensitivity, specificity, ROC?

1. What are precision and recall?

1. Describe t-test in the context of Machine Learning.

## Clustering

1. Describe the k-means algorithm.

1. What is distortion function? Is it convex or non-convex?

1. Tell me about the convergence of the distortion function.

1. Topic: EM algorithm

1. What is the Gaussian Mixture Model?

1. Describe the EM algorithm intuitively.

1. What are the two steps of the EM algorithm

1. Compare GMM vs GDA.

## Dimensionality Reduction

1. Why do we need dimensionality reduction techniques? (data compression, speeds up learning algorithm and visualizing data)

1. What do we need PCA and what does it do? (PCA tries to find a lower dimensional surface such the sum of the squared projection error is minimized)

1. What is the difference between logistic regression and PCA?

1. What are the two pre-processing steps that should be applied before doing PCA? (mean normalization and feature scaling)

## Basics of Natural Language Processing

1. What is WORD2VEC?

1. What is t-SNE? Why do we use PCA instead of t-SNE?

1. What is sampled softmax?

1. Why is it difficult to train a RNN with SGD?

1. How do you tackle the problem of exploding gradients? (By gradient clipping)

1. What is the problem of vanishing gradients? (RNN doesn't tend to remember much things from the past)

1. How do you tackle the problem of vanishing gradients? (By using LSTM)

1. Explain the memory cell of a LSTM. (LSTM allows forgetting of data and using long memory when appropriate.)

1. What type of regularization do one use in LSTM?

1. What is Beam Search?

1. How to automatically caption an image? (CNN + LSTM)

## Miscellaneous

1. What is the difference between loss function, cost function and objective function?

# Deep Learning - List of questions

## Table of Contents

- [General questions](#general-questions)

- [Machine Learning basics](#machine-learning-basics)

- [Optimization procedures](#optimization-procedures)

- [Parameter initialization](#parameter-initialization)

- [Sequence Modeling](#sequence-modeling)

- [Autoencoders](#autoencoders)

- [Representation Learning](#representation-learning)

- [Monte Carlo Methods](#monte-carlo-methods)

- [Adversarial Networks](#adversarial-networks)

## General questions

1. How will you implement dropout during forward and backward pass?

1. What do you do if Neural network training loss/testing loss stays constant? (ask if there could be an error in your code, going deeper, going simpler…)

1. Why do RNNs have a tendency to suffer from exploding/vanishing gradient? How to prevent this? (Talk about LSTM cell which helps the gradient from vanishing, but make sure you know why it does so. Talk about gradient clipping, and discuss whether to clip the gradient element wise, or clip the norm of the gradient.)

1. Do you know GAN, VAE, and memory augmented neural network? Can you talk about it?

1. Does using full batch means that the convergence is always better given unlimited power? (Beautiful explanation by Alex Seewald: https://www.quora.com/Is-full-batch-gradient-descent-with-unlimited-computer-power-always-better-than-mini-batch-gradient-descent)

1. What is the problem with sigmoid during backpropagation? (Very small, between 0.25 and zero.)

1. Given a black box machine learning algorithm that you can’t modify, how could you improve its error? (you can transform the input for example.)

1. How to find the best hyper parameters? (Random search, grid search, Bayesian search (and what it is?))

1. What is transfer learning?

1. Compare and contrast L1-loss vs. L2-loss and L1-regularization vs. L2-regularization.

## Machine Learning basics

1. Can you state Tom Mitchell's definition of learning and discuss T, P and E?

1. What can be different types of tasks encountered in Machine Learning?

1. What are supervised, unsupervised, semi-supervised, self-supervised, multi-instance learning, and reinforcement learning?

1. Loosely how can supervised learning be converted into unsupervised learning and vice-versa?

1. Consider linear regression. What are T, P and E?

1. Derive the normal equation for linear regression.

1. What do you mean by affine transformation? Discuss affine vs. linear transformation.

1. Discuss training error, test error, generalization error, overfitting, and underfitting.

1. Compare representational capacity vs. effective capacity of a model.

1. Discuss VC dimension.

1. What are nonparametric models? What is nonparametric learning?

1. What is an ideal model? What is Bayes error? What is/are the source(s) of Bayes error occur?

1. What is the no free lunch theorem in connection to Machine Learning?

1. What is regularization? Intuitively, what does regularization do during the optimization procedure? (expresses preferences to certain solutions, implicitly and explicitly)

1. What is weight decay? What is it added?

1. What is a hyperparameter? How do you choose which settings are going to be hyperparameters and which are going to be learnt? (either difficult to optimize or not appropriate to learn - learning model capacity by learning the degree of a polynomial or coefficient of the weight decay term always results in choosing the largest capacity until it overfits on the training set)

1. Why is a validation set necessary?

1. What are the different types of cross-validation? When do you use which one?

1. What are point estimation and function estimation in the context of Machine Learning? What is the relation between them?

1. What is the maximal likelihood of a parameter vector $theta$? Where does the log come from?

1. Prove that for linear regression MSE can be derived from maximal likelihood by proper assumptions.

1. Why is maximal likelihood the preferred estimator in ML? (consistency and efficiency)

1. Under what conditions do the maximal likelihood estimator guarantee consistency?

1. What is cross-entropy of loss? (trick question)

## Optimization procedures

1. What is the difference between an optimization problem and a Machine Learning problem?

1. How can a learning problem be converted into an optimization problem?

1. What is empirical risk minimization? Why the term empirical? Why do we rarely use it in the context of deep learning?

1. Name some typical loss functions used for regression. Compare and contrast. (L2-loss, L1-loss, and Huber loss)

1. What is the 0-1 loss function? Why can't the 0-1 loss function or classification error be used as a loss function for optimizing a deep neural network? (Non-convex, gradient is either 0 or undefined. https://davidrosenberg.github.io/ml2015/docs/3a.loss-functions.pdf)

1.

## Parameter initialization

1.

## Sequence Modeling

1. Write the equation describing a dynamical system. Can you unfold it? Now, can you use this to describe a RNN? (include hidden, input, output, etc.)

1. What determines the size of an unfolded graph?

1. What are the advantages of an unfolded graph? (arbitrary sequence length, parameter sharing, and illustrate information flow during forward and backward pass)

1. What does the output of the hidden layer of a RNN at any arbitrary time \_t\_ represent?

1. Are the output of hidden layers of RNNs lossless? If not, why?

1. RNNs are used for various tasks. From a RNNs point of view, what tasks are more demanding than others?

1. Discuss some examples of important design patterns of classical RNNs.

1. Write the equations for a classical RNN where hidden layer has recurrence. How would you define the loss in this case? What problems you might face while training it? (Discuss runtime)

1. What is backpropagation through time? (BPTT)

1. Consider a RNN that has only output to hidden layer recurrence. What are its advantages or disadvantages compared to a RNNhaving only hidden to hidden recurrence?

1. What is Teacher forcing? Compare and contrast with BPTT.

1. What is the disadvantage of using a strict teacher forcing technique? How to solve this?

1.

1. Explain the vanishing/exploding gradient phenomenon for recurrent neural networks. (use scalar and vector input scenarios)

1. Why don't we see the vanishing/exploding gradient phenomenon in feedforward networks? (weights are different in different layers - Random block intialization paper)

1. What is the key difference in architecture of LSTMs/GRUs compared to traditional RNNs? (Additive update instead of multiplicative)

1. What is the difference between LSTM and GRU?

1. Explain Gradient Clipping.

1. Adam and RMSProp adjust the size of gradients based on previously seen gradients. Do they inherently perform gradient clipping? If no, why?

1. Discuss RNNs in the context of Bayesian Machine Learning.

1. Can we do Batch Normalization in RNNs? If not, what is the alternative? (BNorm would need future data; Layer Norm)

## Autoencoders

1. What is an Autoencoder? What does it "auto-encode"?

1. What were Autoencoders traditionally used for? Why there has been a resurgence of Autoencoders for generative modeling?

1. What is recirculation?

1. What loss functions are used for Autoencoders?

1. What is a linear autoencoder? Can it be optimal (lowest training reconstruction error)? If yes, under what conditions?

1. What is the difference between Autoencoders and PCA (can also be used for reconstruction - https://stats.stackexchange.com/questions/229092/how-to-reverse-pca-and-reconstruct-original-variables-from-several-principal-com).

1. What is the impact of the size of the hidden layer in Autoencoders?

1. What is an undercomplete Autoencoder? Why is it typically used for?

1. What is a linear Autoencoder? Discuss it's equivalence with PCA. (only valid for undercomplete) Which one is better in reconstruction?

1. What problems might a nonlinear undercomplete Autoencoder face?

1. What are overcomplete Autoencoders? What problems might they face? Does the scenario change for linear overcomplete autoencoders? (identity function)

1. Discuss the importance of regularization in the context of Autoencoders.

1. Why does generative autoencoders not require regularization?

1. What are sparse autoencoders?

1. What is a denoising autoencoder? What are its advantages? How does it solve the overcomplete problem?

1. What is score matching? Discuss it's connections to DAEs.

1. Are there any connections between Autoencoders and RBMs?

1. What is manifold learning? How are denoising and contractive autoencoders equipped to do manifold learning?

1. What is a contractive autoencoder? Discuss its advantages. How does it solve the overcomplete problem?

1. Why is a contractive autoencoder named so? (intuitive and mathematical)

1. What are the practical issues with CAEs? How to tackle them?

1. What is a stacked autoencoder? What is a deep autoencoder? Compare and contrast.

1. Compare the reconstruction quality of a deep autoencoder vs. PCA.

1. What is predictive sparse decomposition?

1. Discuss some applications of Autoencoders.

## Representation Learning

1. What is representation learning? Why is it useful? (for a particular architecture, for other tasks, etc.)

1. What is the relation between Representation Learning and Deep Learning?

1. What is one-shot and zero-shot learning (Google's NMT)? Give examples.

1. What trade offs does representation learning have to consider?

1. What is greedy layer-wise unsupervised pretraining (GLUP)? Why greedy? Why layer-wise? Why unsupervised? Why pretraining?

1. What were/are the purposes of the above technique? (deep learning problem and initialization)

1. Why does unsupervised pretraining work?

1. When does unsupervised training work? Under which circumstances?

1. Why might unsupervised pretraining act as a regularizer?

1. What is the disadvantage of unsupervised pretraining compared to other forms of unsupervised learning?

1. How do you control the regularizing effect of unsupervised pretraining?

1. How to select the hyperparameters of each stage of GLUP?

1.

## Monte Carlo Methods

1. What are deterministic algorithms? (nothing random)

2. What are Las vegas algorithms? (exact or no solution, random resources)

3. What are deterministic approximate algorithms? (solution is not exact but the error is known)

4. What are Monte Carlo algorithms? (approximate solution with random error)

5.

## Adversarial Networks

1. Discuss state-of-the-art attack and defense techniques for adversarial models.

1.