* You have a cube where there is ½ probability of each vertex being either green or blue. What is the probability that no two adjacent vertices are the same color?
* A baseball player is running towards 3rd base from 2nd at a rate of 20ft/s. He is 15 ft from 3rd base and the baseball diamond has 90ft length for each side. At what rate is he moving towards home base?
* A bank can has an interest rate of 5% for the first 6 months, 9% for the first 9 months. Is the rate 7% for a loan from 6-9 months better?
* Paul ran 3 laps at a rate of 9mph. How fast does he need to for a final 1 lap to average 10mph for the entire trip?
* How many combinations are there for a+b<c if c<=5 and a and b are strictly positive?
* How many digits is 2^40?
* "There are X answers to a certain question on Quora. How do you create model that takes user viewing history to rank the questions? How computationally intensive is this model?"
* "You're drawing from a random variable that is normally distributed X ~ N(0,1), once per day. What is the expected number of days that it takes to draw a value that's higher that 2?"
* <https://www.codechef.com/wiki/tutorial-expectation>
* Expected number of flips to get HHH and then HT
  + <https://math.stackexchange.com/questions/521130/expected-value-of-flips-until-ht-consecutively>
* Scalable ML: What is spark, Hadoop, mapreduce, hive, streaming
* How to build ML workflows (read more into tensorflow)
* LSTM, lookahead bias, bagging, boosting, stacking, information retrieval, batch norm, entity recognition, bootstrapping, Glorot initialization, Kullback-Leibler divergence, GLOVE, SMAPE, HMM, MAP, exponential family, VC dimension, EM, L1, TD(Lambda), instance-based learning, association learning,
* XML, CSS, SQL/SOQL and RegEx
* Suppose you had bank transaction data, and wanted to separate out likely fraudulent transactions. How would you approach it? Why might accuracy be a bad metric for evaluating success?
* What is the equation for Convolutional Neural Networks?
* What is recurrent neural network
* Advantages and disadvantages of neural nets
* What is deep learning
* Tell me the components of a computer and how they work together.
* Map out an example CNN
* Explain Generative Adversarial Networks.
* Given a set of components, what would you do with them to build a system? (The question was purposely vague, so it took awhile to eventually narrow down what the interviewer was trying to get.)
* ROC Curve
* MAP vs MLE and Bayes theorem
* Naïve Bayes
* L1 vs L2 regularization
* **What’s the difference between Type I and Type II error?**
* **What’s a Fourier transform?**
* **What’s the difference between probability and likelihood?**
* **What’s the difference between a generative and discriminative model?**
* **What cross-validation technique would you use on a time series dataset? (**forward chaining where you’ll be able to model on past data then look at forward-facing data.)
* **How is a decision tree pruned?**
* **Which is more important to you– model accuracy, or model performance?**
* **What’s the F1 score? How would you use it?**
* **How would you handle an imbalanced dataset?**
* **When should you use classification over regression?**
* **Name an example where ensemble techniques might be useful.**
* **How do you ensure you’re not overfitting with a model?**
* **What evaluation approaches would you work to gauge the effectiveness of a machine learning model?**
* **What’s the “kernel trick” and how is it useful?**
* **Pick an algorithm. Write the psuedo-code for a parallel implementation.**
* **How would you implement a recommendation system for our company’s users?**
* **How would you simulate the approach AlphaGo took to beat Lee Sidol at Go?**
* *write a function to sample from a multinomial distribution*
* What is the EM algorithm? Give a couple of applications
* What is linear in a generalized linear model?
* What is a probabilistic graphical model? What is the difference between Markov networks and Bayesian networks?
* Give an example of an application of non-negative matrix factorization
* What methods for dimensionality reduction do you know and how do they compare with each other?
* What are some good ways for performing feature selection that do not involve exhaustive search?
* How would you evaluate the quality of the clusters that are generated by a run of K-means?
* What are tensors
* Why population s.d. has N degrees of freedom while sample s.d. has N-1 degrees of freedom? In other words, why 1/N inside root for pop. s.d. and 1/(N-1) inside root for sample s.d.? (Here)
* <https://medium.com/subhrajit-roy/cracking-the-machine-learning-interview-1d8c5bb752d8>
* <https://elitedatascience.com/machine-learning-interview-questions-answers>
* <https://www.edureka.co/blog/interview-questions/machine-learning-interview-questions/>
* <https://career.guru99.com/top-20-mathematics-interview-question/>
* <https://www.themuse.com/advice/7-insane-brain-teasers-you-could-actually-encounter-in-an-interview>
* <https://www.glassdoor.com/Interview/math-interview-questions-SRCH_KT0,4.htm>
* Interpret p-value
  + 0.05 p-value means that assuming the null value is true, we would only expect to get a result this extreme for that statistic 5% of the time
* Interpret confidence interval
  + 95% confidence interval means that for 95% of samples, this method to generate this confidence interval will contain the true population mean 95% of the time
* What is stacking
  + Train a bunch of weak base models on the whole training set and use the predictions as features: [h\_1(x),h\_2(x),…h\_b(x)] which we input for the meta-model to be trained on with the same y label as the label.
* manifold learning
* probabilistic graphical models
* precision and recall
* ANOVA testing
* Central Limit Theorem
* Kalman filters
* <https://www.quora.com/If-you-start-with-1-and-your-friend-with-2-and-the-bet-is-1-until-one-of-you-runs-of-money-What-is-your-chance-of-winning-if-you-have-a-2-3-chance-of-winning-each-bet>