# **Calculus in Data Science and it uses**

A calculus is an abstract theory developed in a purely formal way.

The calculus, more properly called analysis is the branch of mathematics studying the rate of change of quantities (which can be interpreted as slopes of curves) and the length, area, and volume of objects. The calculus is divided into differential and integral calculus.

IMG_257

Differentiation

IMG_258

Integration

*The word Calculus comes from Latin meaning “small stone”,  
Because it is like understanding something by looking at small pieces.*

Calculus is a intrinsic field of maths and especially in many machine learning algorithms that you cannot think of skipping this course to learn the essence of Data Science.

****Differential Calculus**** cuts something into small pieces to find how it changes.

****Integral Calculus**** joins (integrates) the small pieces together to find how much there is.

Now again i would highly recommend you to watch essence of calculus video from [3blue1brown’s channel](https://www.youtube.com/channel/UCYO_jab_esuFRV4b17AJtAw" \t "https://towardsdatascience.com/_blank) which teaches some important pillars of calculus that are required in Data Science.

If you have any kind of allergy or not in the mood of learning through videos, you can refer this. It covers all the basic idea of calculus.

## **[Calculus Menu](https://www.mathsisfun.com/calculus/" \t "https://towardsdatascience.com/_blank)**

### [The word Calculus comes from Latin meaning "small stone", Because it is like understanding something by looking at…](https://www.mathsisfun.com/calculus/" \t "https://towardsdatascience.com/_blank)

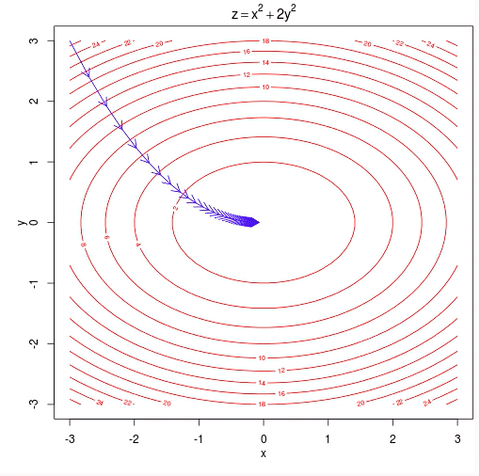
[www.mathsisfun.com](https://www.mathsisfun.com/calculus/" \t "https://towardsdatascience.com/_blank)

I hope you have understood the basics of differentiation and integration. Data Scientists use calculus for almost every model, a basic but very excellent example of calculus in Machine Learning is Gradient Descent.

# **Gradient Descent**

A gradient measures how much the output of a function changes if you change the inputs a little bit.

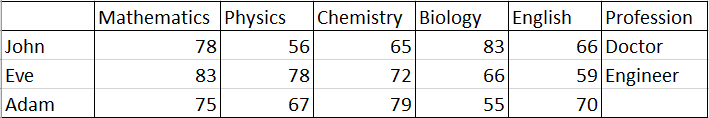
Suppose you have a ball and and a bowl. No matter wherever you slide the ball in the bowl, it will eventually land in the bottom of the bowl.



As you see this ball follows a path that ends at the bottom of the bowl. We can also say that the ball is descending in the bottom of the bowl. As you can see from the image the red lines are gradient of the bowl and the blue line is the path of the ball and as the path of the ball’s slope is decreasing, it is called as gradient descent.

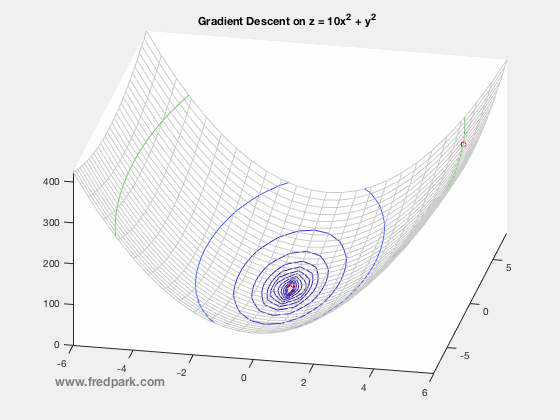
In our machine learning model our goal is to reduce the cost in our input data. the cost function is used to monitor the error in predictions of an ML model. So minimizing this, basically means getting to the lowest error value possible or increasing the accuracy of the model. In short, We increase the accuracy by iterating over a training data set while tweaking the parameters(the weights and biases) of our model.

Let’s us consider we have a dataset of users with their marks in some of the subjects and their occupation. Our goal is to predict the occupation of the person with considering the marks of the person.



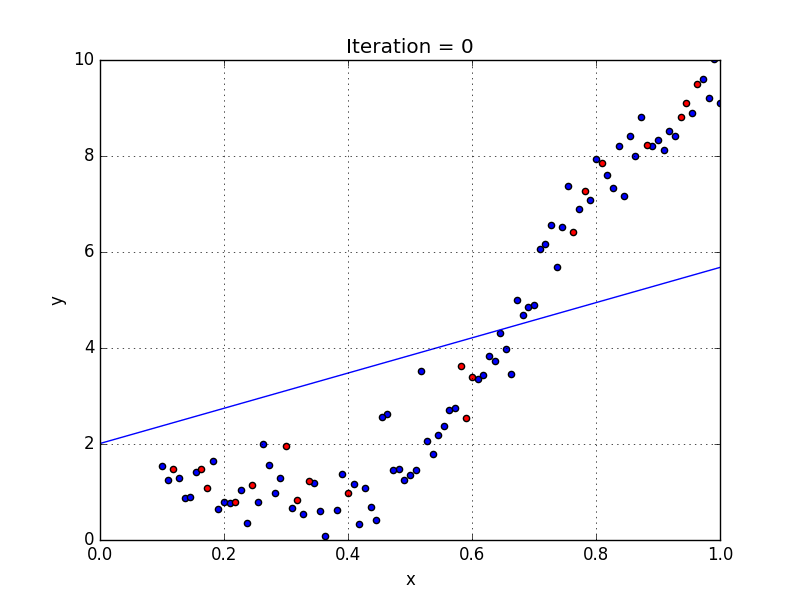
In this dataset we have data of John and eve. With the reference data of john and eve, we have to predict the profession of Adam.

Now think of marks in the subject as a gradient and profession as the bottom target. You have to optimise your model so that the result it predicts at the bottom should be accurate. Using John’s and Eve’s data we will create gradient descent and tune our model such that if we enter the marks of john then it should predict result of Doctor in the bottom of gradient and same for Eve. This is our trained model. Now if we give marks of subject to our model then we can easily predict the profession.



In theory this is it for gradient descent, but to calculate and model, gradient descent requires calculus and now we can see importance of calculus in machine learning.

First Let’s start by the topic that you know till now ie. Linear Algebra. Let first use linear algebra and its formula for our model.



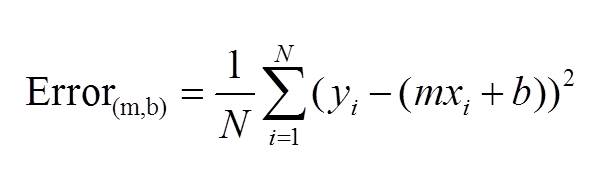
The basic formula that we can use in this model is

y = m\*x +b

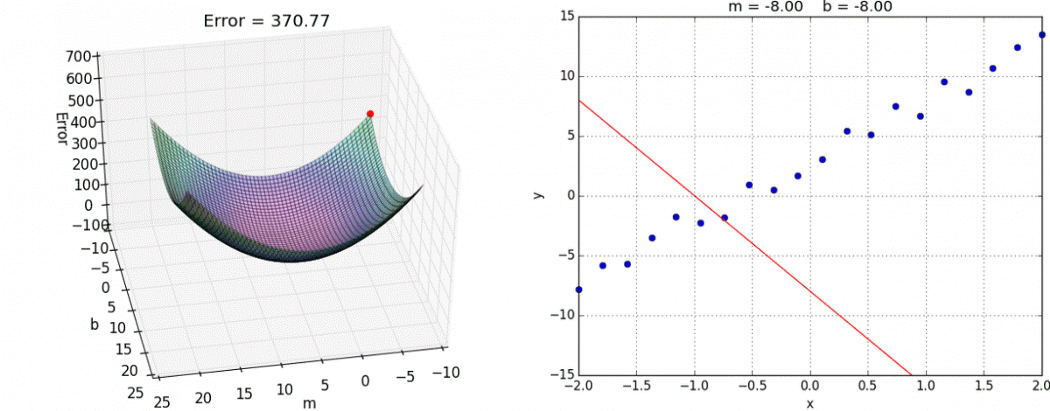
where,

y = predictor, m = slope, x = input, b= y-intercept.

A standard approach to solving this type of problem is to define an error function (also called a cost function) that measures how “good” a given line is. This function will take in a (m,b) pair and return an error value based on how well the line fits our data. To compute this error for a given line, we’ll iterate through each (x,y) point in our data set and sum the square distances between each point’s y value and the candidate line’s y value (computed at mx + b). It’s conventional to square this distance to ensure that it is positive and to make our error function differentiable.

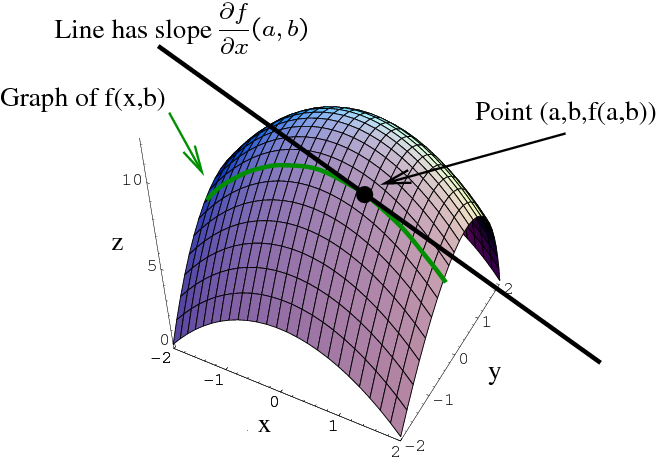


Lines that fit our data better (where better is defined by our error function) will result in lower error values. If we minimize this function, we will get the best line for our data. Since our error function consists of two parameters (m and b) we can visualize it as a two-dimensional surface. This is what it looks like for our data set:



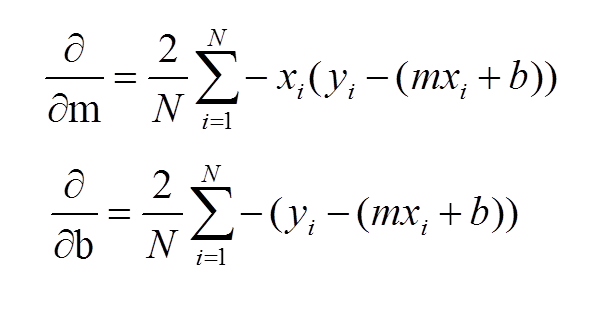
Each point in this two-dimensional space represents a line. The height of the function at each point is the error value for that line. You can see that some lines yield smaller error values than others (i.e., fit our data better). When we run gradient descent search, we will start from some location on this surface and move downhill to find the line with the lowest error.

In the essence of calculus video you have seen that to calculate slope, we use differentiation.



The graph of a function z=f(x,y)z=f(x,y) is a surface, and fixing y=by=b gives a curve (shown in green). The partial derivative ∂f∂x(a,b)∂f∂x(a,b) is the slope of the tangent line to this curve at the point where x=ax=a.

To run gradient descent on this error function, we first need to compute its gradient. The gradient will act like a compass and always point us downhill. To compute it, we will need to differentiate our error function. Since our function is defined by two parameters (m and b), we will need to compute a partial derivative for each. These derivatives work out to be:

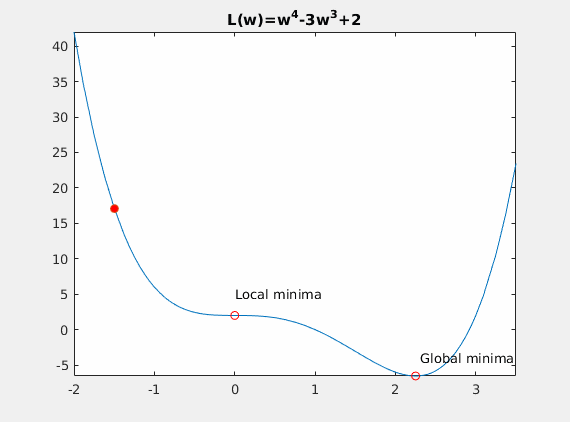


We now have all the tools needed to run gradient descent. We can initialize our search to start at any pair of m and b values (i.e., any line) and let the gradient descent algorithm march downhill on our error function towards the best line. Each iteration will update m and b to a line that yields slightly lower error than the previous iteration. The direction to move in for each iteration is calculated using the two partial derivatives from above.

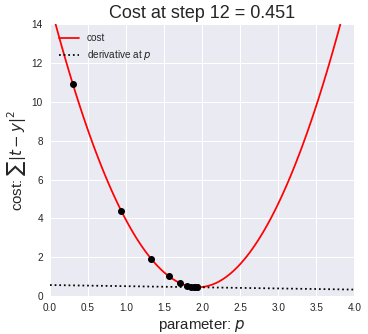
The****Learning Rate**** variable controls how large of a step we take downhill during each iteration. If we take too large of a step, we may step over the minimum. However, if we take small steps, it will require many iterations to arrive at the minimum.

While we were able to scratch the surface for learning gradient descent, there are several additional concepts that are good to be aware of that we weren’t able to discuss. A few of these include:

* ****Convexity**** – In our linear regression problem, there was only one minimum. Our error surface was [convex](http://en.wikipedia.org/wiki/Convex_function" \t "https://towardsdatascience.com/_blank). Regardless of where we started, we would eventually arrive at the absolute minimum. In general, this need not be the case. It’s possible to have a problem with local minima that a gradient search can get stuck in. There are several approaches to mitigate this (e.g., [stochastic gradient search](http://en.wikipedia.org/wiki/Stochastic_gradient_descent" \t "https://towardsdatascience.com/_blank)).



* ****Convergence**** – We didn’t talk about how to determine when the search finds a solution. This is typically done by looking for small changes in error iteration-to-iteration (e.g., where the gradient is near zero).



# **Multivariate Calculus**

Now lets deep dive to Multivariable calculus which will teach calculus in multivariable data which we will get eventually in the real life.

## **[Multivariable Calculus | Khan Academy](https://www.khanacademy.org/math/multivariable-calculus" \t "https://towardsdatascience.com/_blank)**

### [Learn for free about math, art, computer programming, economics, physics, chemistry, biology, medicine, finance…](https://www.khanacademy.org/math/multivariable-calculus" \t "https://towardsdatascience.com/_blank)

[www.khanacademy.org](https://www.khanacademy.org/math/multivariable-calculus" \t "https://towardsdatascience.com/_blank)

*To get the latest updates, tips and anything you want or have issue just post in the comments.*

### **1. Describe the various branches of mathematics with which you are familiar. Which one do you prefer?**

Employers may ask you about the types of math you know and feel comfortable with, so they can assess your strengths and determine if you possess the necessary skills for the job. After briefly listing a few math areas you have experience with, describe the branch of math you are proficient in.

**Example:** *"My math skills include basic arithmetic, algebra, geometry and calculus. My interest in calculus comes from my ability to understand how parameters change with respect to one another. I can appreciate the intricacies of complex systems, how variables are interrelated and how a change in one can affect the other. The other area that interests me is statistics. It helps me understand the data better and appreciate how professionals processed large amounts of data to design complex systems."*

**Related:** **[How To Use The STAR Interview Response Technique](https://in.indeed.com/career-advice/interviewing/how-to-use-the-star-interview-response-technique" \t "https://in.indeed.com/career-advice/interviewing/_blank)**

### **2. What mathematics courses have you taken?**

The interviewer may ask this question so they can find out how much formal math training you have, the courses you have taken and your understanding of various concepts. Explain to the interviewer how and where you used the concepts you studied.

**Example:** *"I took up geometry, algebra and trigonometry in high school. At college, I attended courses that included probability, statistics, linear algebra, calculus and vectors. I was able to use these concepts in my computer science courses to design different algorithms and machine learning models."*

**Related:** **[Skills Test: Definition And Examples](https://in.indeed.com/career-advice/career-development/skills-test" \t "https://in.indeed.com/career-advice/interviewing/_blank)**

### **3. What are some scenarios where you put your math concepts to use?**

This question enables the interviewer to see if you have applied your knowledge and skills to solving a real-world problem. It helps them understand your skill-set and your ability to apply it in specific scenarios to get the desired results.

**Example:** "*I used my knowledge in statistics to perform exploratory data analysis on a school dataset. I was able to obtain information on the gender ratio, the salaries of teachers, the number of years of experience, the grades that students attain in different subjects and the number of admissions each academic year from the database. The result of this was a more in-depth analysis of the growth of a school over time and its areas for improvement.*"

**Related:** **[Communication Skills: Definitions And Examples](https://in.indeed.com/career-advice/resumes-cover-letters/communication-skills" \t "https://in.indeed.com/career-advice/interviewing/_blank)**

### **4. What is the central limit theorem?**

In addition to determining whether you are comfortable with this kind of math, a hiring manager can assess how confident you are in the concepts that you know. Furthermore, they may want to look at your ability to communicate about a topic and your in-depth understanding of it.

**Example:** *"When you take large random samples with replacement from a population with specific mean and standard deviation, the sample mean distribution corresponds to a normal distribution. Even if the distribution of the data is not normal, the distribution of the means of the samples drawn from it would be normal. It can be used to estimate the average family income in a region."*

### **5. Let A and B be two events in the same sample space, with P (A) = 0.4 and P (B) = 0.8. Can you call them disjoint events?**

Potential employers might ask this question to determine your level of comfort with basic math skills and how strong your fundamental concepts are. It is essential to be concise and clear. You can provide an example if possible.

**Example:** *"Events that are disjoint never occur at the same time. In the above questions, the two events cannot be disjoint. If they are disjoint then the total probability would be 0.4 + 0.8 which is 1.2. This is contrary to the law that probability can never exceed 1."*

**Related\*\***:**[**Top 15 Careers in Mathematics (And Salary Information)\*\*]([https://in.indeed.com/career-advice/finding-a-job/career-in-mathematics](https://in.indeed.com/career-advice/finding-a-job/career-in-mathematics" \t "https://in.indeed.com/career-advice/interviewing/_blank))

### **6. Explain the concept of Bayes' theorem?**

During the interview, you can expect similar questions. Provide an example to demonstrate your understanding. The interviewer can see that you have a strong grasp of fundamental concepts.

**Example:** *"The Bayes' theorem describes how likely an event is to occur given any condition. Bayes' theorem can be used to calculate conditional probability. As an example: if you give four bags of balls, each with three balls of varying colours: red, green and blue and we calculate the probability of taking a green ball from the fourth bag. When a probability of an event depends on other factors, we call it conditional probability. We can use the formula of Bayes' theorem to calculate the probability of this event."*

### **7. What is a unit vector?**

The interviewer may proceed to ask questions from other areas of mathematics. The interviewer can determine what your strengths and weaknesses are in the other areas.

**Example:** *"Vectors are quantities with both magnitude and direction. Unit vectors have magnitudes of 1. Dividing a vector by its magnitude makes it a unit vector. Unit vectors are commonly used to indicate direction and the scalar coefficients provide magnitude."*

### **8. What is a position vector? Where do you use them?**

If the interviewer wants to gauge your understanding of a topic further, they may ask a similar question as a follow-up question. Whenever possible, consider adding examples; this shows your grasp of the topic and ability to explain it through analogies.

**Example:** "*A position vector specifies the location of a body. To describe the motion of a body, it is vital to understand its position. Position vectors change length or direction or both as the point moves. Thus, they can be used to track the current position of a particular point in space."*

### **9. What is the rank of the matrix? What is the rank of a null matrix?**

There can be theoretical questions that assess your ability to recall concepts from various topics and use your conceptual abilities to answer them. Explain how you arrived at each step and how you made your conclusion.

**Example:** *"The rank of a matrix is the maximum number of linearly independent columns or rows of a matrix. The rank cannot exceed the number of row or column elements. There are no rows or columns in a null matrix. Therefore, this matrix has a rank of zero."*

### **10. What is a linear equation? Where can you use it?**

Such questions are a good way of assessing your capability to apply a concept in a real-life situation. Provide examples that clearly demonstrate how you applied these concepts to solve a real-world problem.

**Example:** "*L\*\*inear equations are equations of the first order. We call equations as linear equations in one variable if they contain only one variable. There can be more than one variable in a linear equation. In that case, we call them linear equations in two variables and so on. A linear equation can be used to represent almost any scenario where an unknown quantity is present, such as determining income over time, calculating speed or distance and predicting profit."*

### **11. What is the limit of a function?**

If you mentioned any areas of interest at the start of the interview, the interviewer can ask questions related to those. By asking these questions, the interviewer can determine your strengths and your approach towards these.

**Example:** *"A function's limit at a point a in its domain is the value the function approaches as its argument approaches a. In simple words, the limit is the value that a function approaches as inputs to that function approaches a specified number."*

## **Background Questions**

Usually, interviews start with background questions: they can ask you to talk about yourself. This can also happen at the telephone interview stage.

For background questions be ready to talk about a summary of your career.

* Summarize your experience
* What companies you worked at? What was your role?
* Do you have a project portfolio? What projects you implemented? Discuss some of them in details
* For graduating students: Tell me about your master thesis
* For aspiring data scientists: Why do you want a career in data science?
* Have you taken any data-science-related online courses? If yes, how many did you complete with a certificate?
* Have you participated in any data science challenges? If yes, can you describe one of them?

## **Process**

All Machine Learning, Data Mining and Data Science projects should follow some process, so there can be questions about it:

* Can you outline the steps in a data science project?
* Have you heard of CRISP-DM (Cross Industry Standard Process for Data Mining)?

CRISP-DM defines the following steps:

* Problem Definition
* Data Understanding (or Data Exploration)
* Data Preparation
* Modeling
* Evaluation
* Deployment (for the production)

So next you may discuss each of these steps in details

* What is the goal of each step?
* What are possible activities at each step?

## **Mathematics**

Some background mathematics is necessary for doing Data Science, therefore you may expect math-related questions. On the other hand, for some Data Science positions there could be very few math questions, or none at all. In my opinion, it's always better to know the underlying theory when talking about Machine Learning algorithms, but your interviewers may have a different point of view.

### Linear Algebra

Basic Linear Algebra questions might include:

* What is Ax=bAx=b? How to solve it?
* How do we multiply matrices?
* What is an Eigenvalue? And what is an Eigenvector? What is Eigenvalue Decomposition or The Spectral Theorem?
* What is Singular Value Decomposition?
* You may expect Liner Algebra questions in the Machine Learning part of the interview (see below).

If you are interested in learning or refreshing Linear Algebra, see [Best Time to Learn Linear Algebra is Now!](http://www.itshared.org/2015/02/best-time-to-learn-linear-algebra-is-now.html)

### Other Areas

* Discrete Mathematics and Logics are not that important for Data Science
* Probability and Statistics are core skills and discussed in the next section
* Calculus and Optimization are usually discussed in the Machine Learning part and usually when talking about a particular algorithm

## **Probability and Statistics**

Probability and Statistics give the foundation for Machine Learning, which makes them an important subject. It also may be useful if the company is doing some marketing or website optimization, so they could ask about related concepts such as A/B tests.

### Basic Probability

You can have a couple of simple questions to check your understanding of probability.

For example:

* Given two fair dices, what is the probability of getting scores that sum to 4? to 8?
* A simple questions on Bayes rule: Imagine a test with a true positive rate of 100% and false positive rate of 5%. Imagine a population with a 1/1000 rate of having the condition the test identifies. Given a positive test, what is the probability of having that condition?

### Distributions

You may expect questions about probability distributions:

* What is the normal distribution? Give an example of some variable that follows this distribution
* What about log-normal?
* Explain what a long tailed distribution is and provide three examples of relevant phenomena that have long tails. Why are they important in classification and prediction problems?
* How to check if a distribution is close to Normal? Why would you want to check it? What is a QQ Plot?
* Give examples of data that does not have a Gaussian distribution, or log-normal.
* Do you know what the exponential family is?
* Do you know the Dirichlet distribution? the multinomial distribution?

### Basic Statistics

* What is the Laws of Large Numbers? Central Limit Theorem?
* Why are they important for Statistics?
* What summary statistics do you know?

### Experiment Design

Designing experiments is an important part of Statistics, and it’s especially useful for doing A/B tests.

Sampling and Randomization

* Why do we need to sample and how?
* Why is randomization important in experimental design?
* Some 3rd party organization randomly assigned people to control and experiment groups. How can you verify that the assignment truly was random?
* How do you calculate needed sample size?
* Power analysis. What is it?

Biases

* When you sample, what bias are you inflicting?
* How do you control for biases?
* What are some of the first things that come to mind when I do X in terms of biasing your data?

Other questions

* What are confounding variables?

### Point Estimates

Confidence intervals

* What is a point estimate? What is a confidence interval for it?
* How are they constructed?
* How to interpret confidence intervals?

### Testing

Hypothesis tests

* Why do we need hypothesis testing? What is P-Value?
* What is the null hypothesis? How do we state it?
* Do you know what Type-I/Type-II errors are?
* What is tt-Test/FF-Test/ANOVA? When to use it?
* How would you test if two populations have the same mean? What if you have 3 or 4 populations?
* You applied ANOVA and it says that the means are different. How do you identify the populations where the differences are significant?
* What is the distribution of p-value’s, in general?

### A/B Tests

* What is A/B testing? How is it different from usual Hypothesis testing?
* How can you prove that one improvement you’ve brought to an algorithm is really an improvement over not doing anything? How familiar are you with A/B testing?
* How can we tell whether our website is improving?
* What are the metrics to evaluate a website? A search engine?
* What kind of metrics would you track for you music streaming website?
* Common metrics: Engagement / retention rate, conversion, similar products / duplicates matching, how to measure them.
* Real-life numbers and intuition: Expected user behavior, reasonable ranges for user signup / retention rate, session length / count, registered / unregistered users, deep / top-level engagement, spam rate, complaint rate, ads efficiency.

### Time Series

* What is a time series?
* Did you do any projects which involved dealing with time?
* What is the difference between data for usual statistical analysis and time series data?
* Have you used any of the following: Time series models, Cross-correlations with time lags, Correlograms, Spectral analysis, Signal processing and filtering techniques? If yes, in which context?
* In time series modeling how can we deal with multiple types of seasonality like weekly and yearly seasonality?

### Advanced

Resampling

* Explain what resampling methods are. Why they are useful. What are their limitations?
* Bootstrapping - how and why it is used?
* How to use resampling for hypothesis testing? Have you heard of Permutation Tests?
* How would you apply resampling to time series data?

## **Machine Learning**

In my experience, the Machine Learning part is usually the largest part of the interview. It may be a few basic questions, but it’s helpful to be prepared to more in-depth Machine Learning questions, especially if you claim to have worked with it on your CV.

### General ML Questions

The ML part may start with something like:

* What is the difference between supervised and unsupervised learning? Which algorithms are supervised learning and which are not? Why?
* What is your favorite ML algorithm and why?

And then go into details

### Regression

* Describe the regression problem. Is it supervised learning? Why?
* What is linear regression? Why is it called linear?
* Discuss the bias-variance tradeoff.

Linear Regression:

* What is Ordinary Least Squares Regression? How it can be learned?
* Can you derive the OLS Regression formula? (For one-step solution)
* Is model Y∼X1+X2+X1X2Y∼X1+X2+X1X2 still linear? Why?
* Do we always need the intercept term? When do we need it and when do we not?
* What is collinearity and what to do with it? How to remove multicollinearity?
* What if the design matrix is not full rank?
* What is overfitting a regression model? What are ways to avoid it?
* What is Ridge Regression? How is it different from OLS Regression? Why do we need it?
* What is Lasso regression? How is it different from OLS and Ridge?

Linear Regression assumptions:

* What are the assumptions required for linear regression?
* What if some of these assumptions are violated?

Significant features in Regression

* You would like to find significant features. How would you do that?
* You fit a multiple regression to examine the effect of a particular feature. The feature comes back insignificant, but you believe it is significant. Why can it happen?
* Your model considers the feature XX significant, and ZZ is not, but you expected the opposite result. Why can it happen?

Evaluation

* How to check is the regression model fits the data well?

Other algorithms for regression

* Decision trees for regression
* kk-Nearest Neighbors for regression. When to use?
* Do you know others? E.g. Splines? LOESS/LOWESS?

### Classification

Basic:

* Can you describe what is the classification problem?
* What is the simplest classification algorithm?
* What classification algorithms do you know? Which one you like the most?

Decision trees:

* What is a decision tree?
* What are some business reasons you might want to use a decision tree model?
* How do you build it? What impurity measures do you know?
* Describe some of the different splitting rules used by different decision tree algorithms.
* Is a big brushy tree always good? Why would you want to prune it?
* Is it a good idea to combine multiple trees?
* What is Random Forest? Why is it good?
* Other ways to combine trees? What about boosting?

Logistic regression:

* What is logistic regression?
* How do we train a logistic regression model?
* How do we interpret its coefficients?

Support Vector Machines

* What is the maximal margin classifier? How this margin can be achieved and why is it beneficial?
* How do we train SVM? What about hard SVM and soft SVM?
* What is a kernel? What's the intuition behind the Kernel trick?
* Which kernels do you know? How to choose a kernel?

Neural Networks

* What is an Artificial Neural Network?
* How to train an ANN? What is back propagation?
* How does a neural network with three layers (one input layer, one inner layer and one output layer) compare to a logistic regression?
* What is deep learning? What is CNN (Convolution Neural Network) or RNN (Recurrent Neural Network)?

Other models:

* What other models do you know?
* How can we use Naive Bayes classifier for categorical features? What if some features are numerical?
* Tradeoffs between different types of classification models. How to choose the best one?
* Compare logistic regression with decision trees and neural networks.

### Regularization

* What is Regularization?
* Which problem does Regularization try to solve?
* What does it mean (practically) for a design matrix to be “ill-conditioned”?
* When might you want to use ridge regression instead of traditional linear regression?
* What is the difference between the L1L1 and L2L2 regularization?
* Why (geometrically) does LASSO produce solutions with zero-valued coefficients (as opposed to ridge)?
* Let us go through the derivation of OLS or Logistic Regression. What happens when we add L2L2 regularization? How do the derivations change? What if we replace L2L2 regularization with L1L1 regularization?

### Dimensionality Reduction

Basics:

* What is the purpose of dimensionality reduction and why do we need it?
* Are dimensionality reduction techniques supervised or not? Are all of them are (un)supervised?
* What ways of reducing dimensionality do you know?
* Is feature selection a dimensionality reduction technique?
* What is the difference between feature selection and feature extraction?
* Is it beneficial to perform dimensionality reduction before fitting an SVM? Why or why not?

Principal Component Analysis:

* What is Principal Component Analysis (PCA)? What is the problem it solves? How is it related to eigenvalue decomposition (EVD)?
* What’s the relationship between PCA and SVD? When SVD is better than EVD for PCA?
* Under what conditions is PCA effective?
* Why do we need to center data for PCA and what can happed if we don’t do it? Do we need to scale data for PCA?
* Is PCA a linear model or not? Why?

Other Dimensionality Reduction techniques:

* Do you know other Dimensionality Reduction techniques?
* What is Independent Component Analysis (ICA)? What’s the difference between ICA and PCA?
* Suppose you have a very sparse matrix where rows are highly dimensional. You project these rows on a random vector of relatively small dimensionality. Is it a valid dimensionality reduction technique or not?
* Have you heard of Kernel PCA or other non-linear dimensionality reduction techniques? What about LLE (Locally Linear Embedding) or tt-SNE (tt-distributed Stochastic Neighbor Embedding)
* What is Fisher Discriminant Analysis? How it is different from PCA? Is it supervised or not?

### Cluster Analysis

* What is the cluster analysis problem?
* Which cluster analysis methods you know?
* Describe KK-Means. What is the objective of KK-Means? Can you describe the Lloyd algorithm?
* How do you select KK for K-Means?
* How can you modify KK-Means to produce soft class assignments?
* How to assess the quality of clustering?
* Describe any other cluster analysis method. E.g. DBSCAN.

### Optimization

You may have some basic questions about optimization:

* What is the difference between a convex function and non-convex?
* What is Gradient Descent Method?
* Will Gradient Descent methods always converge to the same point?
* What is a local optimum?
* Is it always bad to have local optima?

### Recommendation

* What is a recommendation engine? How does it work?
* Do you know about the Netflix Prize problem? How would you approach it?
* How to do customer recommendation?
* What is Collaborative Filtering?
* How would you generate related searches for a search engine?
* How would you suggest followers on Twitter?

### Feature Engineering

* How to apply Machine Learning to audio data, images, texts, graphs, etc?
* What is Feature Engineering? Can you give an example? Why do we need it?
* How to go from categorical variables to numerical?
* What to do with categorical variables of high cardinality?

### Natural Language Processing

If the company deals with text data, you can expect some questions on NLP and Information Retrieval:

* What is NLP? How is it related to Machine Learning?
* How would you turn unstructured text data into structured data usable for ML models?
* What is the Vector Space Model?
* What is TF-IDF?
* Which distances and similarity measures can we use to compare documents? What is cosine similarity?
* Why do we remove stop words? When do we not remove them?
* Language Models. What is NN-Grams?
* What is word2vec? How it can be used in NLP and IR?

### Meta Learning

Feature Selection:

* Are all features equally good?
* What are the downfalls of using too many or too few variables?
* How many features should you use? How do you select the best features?
* What is Feature Selection and why do we need it?
* Describe several feature selection methods. Are these methods depend on the model or not?

Model selection:

* You have built several different models. How would you select the best one?
* You have one model and want to find the best set of parameters for this model. How would you do that?
* How would you look for the best parameters? Do you know something else apart from grid search?
* What is Cross-Validation?
* What is 10-Fold CV?
* What is the difference between holding out a validation set and doing 10-Fold CV?

Model evaluation

* How do you know if your model overfits?
* How do you assess the results of a logistic regression?
* Which evaluation metrics you know? Something apart from accuracy?
* Which is better: Too many false positives or too many false negatives?
* What precision and recall are?
* What is a ROC curve? What is AU ROC (AUC)? How to interpret the curve and AU ROC?
* Do you know about Concordance or Lift?

Discussion Questions:

* You have a marketing campaign and you want to send emails to users. You developed a model for predicting if a user will reply or not. How can you evaluate this model? Is there a chart you can use?

### Miscellanea

Curse of Dimensionality

* What is Curse of Dimensionality? How does it affect distance and similarity measures?
* What are the problems of large feature space? How does it affect different models, e.g. OLS? What about computational complexity?
* What dimensionality reductions can be used for preprocessing the data?
* What is the difference between density-sparse data and dimensionally-sparse data?

Others

* You are training an image classifier with limited data. What are some ways you can augment your dataset?

## **Computer Science**

Knowledge in Computer Science is as important for Data Science as knowledge in Machine Learning. So you may get the same type of questions as for any software developer position, but possibly with lower expectations on your answers.

I was a Java developer for quite some time, and I prepared a list of questions I asked (and often was asked) on Java interviews: [Java Inteview questions](http://www.itshared.org/2015/09/java-interview-questions.html). This list can also be helpful for preparing to a Data Science interview.

### Libraries and Tools

Apart from basics of Java/Scala/Python/etc, you may be asked about libraries for data analysis:

* Which libraries for data analysis do you know in Python/R/Java?
* Have you used numpy, scipy, pandas, sklearn?
* What are some features of the sklearn api that differentiate it from fitting models in R?
* What are some features of pandas/sklearn that you like? Don't like? Same questions for R.
* Why is “vectorization” such a powerful method for optimizing numerical code? What is going on that makes the code faster relative to alternatives like nested for loops?
* When is it better to write your own code than using a data science software package?
* State any 3 positive and negative aspects about your favorite statistical software.
* Describe a difficult bug you’ve encountered and how you resolved it.
* How does floating point affect precision of calculations? Equality tests?
* What is BLAS? LAPACK?

### Databases

* Have you been involved in database design and data modeling?
* SQL-Related questions: e.g. what is "group by"?
* Or given some DB schema you may be asked to write a simple SQL query.
* What is a “star schema”? “snowflake schema”?
* Describe different NoSQL technologies you’re familiar with, what they are good at, and what they are bad at.

### Distributed Systems and Big Data

Basic “Big Data” questions:

* What is the biggest data set that you have processed and how did you process it? What was the result?
* Have you used Apache Hadoop, Apache Spark, Apache Flink? Why? Have you used Apache Mahout?

MapReduce

* What is MapReduce? Why is it “shared-nothing” architecture?
* Can you implement word count in MapReduce? What about something a bit more complex like TF-IDF? Naive Bayes?
* What is load balance? How to make sure a MapReduce application has good load balance?
* Can you give examples where MapReduce does not work?
* What are examples of “embarassingly parallelizable” algorithms?
* How would you estimate the median of a dataset that is too big to hold in the memory?

Implementation questions

There are some posts that you may find useful when preparing for the “Big Data” part:

* [Hadoop and MapReduce](http://www.itshared.org/2015/03/hadoop-and-mapreduce.html)
* [Naive Bayes on Apache Flink](http://www.itshared.org/2015/03/naive-bayes-on-apache-flink.html)

## **Hands-On**

Also, many interviews have a part which I call “hands-on”: you are given some problem description and you are asked to solve it. You can just talk the interviewers through your solution or even be asked to sit and implement some parts. Sometimes there is also a test assignment to be done at home (prior to the interview).

### Problem to Solve

For example:

Assume that you are asked to lead a project on churn detection, and have dataset of known users who stopped using the service and ones who are still using. This data includes demographics and other features.

Do the following:

1. Describe the methodology and model that you will chose to identify churn, and describe your thought process.
2. Think how would you communicate the results to the CEO?
3. Suppose in the dataset only 0.025 of users churned. How would you make it more balanced?

Also:

* How would you implement it if you had one day? One month? One year?
* How would your approach scale?

Other problems:

* How would you approach identifying plagiarism?
* How to find individual paid accounts shared by multiple users?
* How to detect bogus reviews, or bogus Facebook accounts used for bad purposes?
* Usually the domain of the problem is related to what the company is doing. If they’re doing marketing, it will most likely be marketing related.

Additionally, you may be asked:

* How would you approach collecting the data if you didn’t have the dataset?

### Coding

Sometimes you even may be presented a small dataset and ask to do a particular task with any tool. For example,

* write a script to extract features,
* then do some exploratory data analysis and
* finally apply some ML algorithm to this dataset.

Or just the last two, with a ready to use dataset in tabular form.

## ****1.    Could You Please Tell Us Some of The Classes of  Math That You Are Familiar With? Which Ones Do You Prefer?****

 I have interacted with lots of Math in my career. However, I mostly come across simple arithmetic, algebra and calculus. The first one comes in handy when counting back change, whereas algebra helps me figure out an unknown variable, such as the number of clients. I usually use calculus for relatively challenging problems, such as calculating the rate of change.

## ****2.    Take Us Through Some of The Math Courses That You Have Taken****

My history with Math dates back to high school. Our curriculum was drafted such that we all took a course in Geometry, Algebra and Trigonometry. However, I also pursued Calculus as an offering. I later advanced in college and pursued further Calculus as part of my degree, which greatly expanded my knowledge.  I use all the skills I learnt in my everyday activities to stay on top of my game.

## ****3.    Walk Us Through the Process of Calculating Interest****

Different kinds of interest can be calculated depending on the situation. An account may either earn simple or compound interest. To calculate simple interest, the percentage of principal is added to the base over each period. For compound interest, the previously earned interest is included for every forthcoming period. My career as a financial analyst saw me do most of these calculations.

## ****4.    What Do You Understand by The Term Percentage? Have You Ever Used These Percentages As Work?****

A percentage is a number or portion out of 100. They play an essential role in different work applications, such as conducting market research and various calculations. My last job as a Mathematics teacher required me to perform lots of classroom research to determine the number of students who observed deadlines and compare the number to the end term scores.

## ****5.    What Is the Mean and How Is It Calculated?****

The mean is the average of a set of numbers. To arrive at it, add all the numbers and divide the final value by the number of elements in the list. If my list contains 4,7,8, and 9, I will add all of them to obtain 28 and then divide them by 4 to get a mean of 7. The mean plays a vital role in the work environment as it can be used to figure out the staffing levels, customer flows and average tips.

## ****6.    Walk Us Through How to Convert a Fraction to a Percentage and Vice Versa****

The easiest way to convert a fraction into a percentage is to divide the top number by the bottom and multiply by 100. If you would like to convert 1/5 into a percentage, divide one by five and multiply by 100 to get 20%.

Turning a percentage into a fraction is also easy enough. You only need to place the percentage over 100 and break down your answer. If you want to turn the 20% into a fraction, place it over 100 and then find a standard number that can divide both the top and bottom values till you arrive at 1/5.

## ****7.    How Can You Make Someone to Be Interested in Math?****

Whenever I engage someone who does not enjoy Math, I use real-world situations to make them curious. I also take some time to find out more about their interests and incorporate them into our discussions or lessons. I once dealt with a student who did not like Math. I made it a norm to use practical examples to ignite his interest and help him understand.

## ****8.    Take Us Through Discount Estimation. How Do You Do It?****

When estimating discounts, I work in multiples of say 5 or 10 instead of finding out the exact percentage. For example, if the discount to be determined is 18%, I work on it with the notion that the value is close to 20%. Therefore, to figure out 10%, I divide the total by ten and then multiply it by 2 to get an approximate discount.

## ****9.    Take Us Through How You Count Back Change for a Customer****

The best way of counting change for a customer is by calculating the total amount of purchase and then the amount of cash they have offered you. For example, if a customer spends $8.75 and offers you a ten-dollar bill, first outline the value of purchase, i.e. $ 8.75, before adding quarters until you get a total dollar amount. Repeat this same process until you reach the amount of cash that the customer has offered.

****[VIDEO] Top 25 Mathematics Interview Questions with Sample Answers: ►**[Subscribe for more useful videos](https://www.youtube.com/projectpractical?sub_confirmation=1" \t "https://www.projectpractical.com/top-25-mathematics-interview-questions-and-answers/_blank)**

## ****10. What Is Algebra?****

Algebra is a field of math that uses symbols and letters to stand in for numbers, points and objects, as well as detail the relationship between them. It is used when determining the unknown variable. If I have lost six pineapples and is left with seven pineapples now, I will use algebra to establish the total number of pineapples I had.

## ****11. What Do You Understand by Commutative, Associative and Distributive Laws?****

In communicative laws, one can swap numbers and still get the same value after addi9ng them. For example, c+b = b+c. In associative laws, the final value remains the same regardless of how the numbers are grouped. For example, (c+b) +d= (c+b) + d. Finally, for distributive laws, we can get the same answer when we multiply a number by a group of numbers added together or by multiplying them separately and then adding them, i.e. bc(d+c) = bcd+ bcc

## ****12. What Is Geometry?****

Geometry is one of the branches of Math. It is divided into two. The first one is plane Geometry, which deals with flat shapes such as triangles, lines and circles, which can be easily drawn on a piece of paper. On the other hand, Solid Geometry focuses on 3D objects, such as cubes, prisms, spheres, and cylinders.

## ****13. Could You Please Define a Linear Equation and What It Is Used For? Also, Please Give Us An Example of How You Can Use It In Your Daily Life.****

A linear equation is a construction used to solve problems involving distance, time and speed. It can also be used to determine weight, mass and density. It is expressed as Bx+ Cy+ …= D. for example, if your home is 30km and you need to reach the office by 8 am, at a speed of 50km/h, you can use a linear equation to know when you should leave home.

## ****14. How Many Doors Are in This Neighborhood?****

To calculate the number of doors in this neighbourhood, I will need you to give me more information. One, I’d like to know your understanding of a door, for example, do I only concentrate on the doors on buildings and offices or others such as car doors? I’d also need the approximate number of doors found in each item and the number of items or buildings present in the neighbourhood. Once I have all that information, calculating the total number of doors in the area will be easy.

## ****15. Could You Please Tell Us the Formula of Calculating Interest?****

There are two different formulas depending on the interest one is calculating. You can either calculate simple or compound interest. For simple interest, the formula that applies is (P \* R\* T)/100, where P stands for Principal, R the rate of interest and T, time. The formula for compound interest is P(1+r/100) t.

## **16. **How Do You Calculate Complex Equations to Arrive at An Accurate Answer?****

To avoid unnecessary confusion, I usually use BODMAS to solve any complex calculation that comes my way. The B stands for Brackets, O for orders, including powers and square roots, DM for division and multiplication, and finally AS for addition and subtraction. Once I am done with the B or O, I proceed from left to right, depending on the problem. The same order applies even to the most technical of problems.

## **17. **Please Tell What You Understand by Tangent and Standard Deviation?****

A tangent is a line that touches a curve at one given point but does not pass or cut across it. On the other hand, a standard deviation is the measure of the spread out of the data about the mean value. It is usually referred to as the sigma and has its definitive symbol.

## **18.  **Why Do You Think You Are the Perfect Fit for This Job?****

I am passionate about calculations aimed at solving problems. I have vast experience, having come across several Mathematical situations in the course of my career. I have won several awards as an accountant and fiscal analyst, showing just how good I am. Therefore, I can assure you that I will use all these experiences to better this organization. I am also a diligent worker who is willing to perform if given a chance.

## ****19. Could You Please Differentiate a Line, Point, Plane and Solid?****

The difference between the three exists in the dimensions. A point has no dimension, whereas a line is one dimensional. A plane has two dimensions, whereas a solid is three dimensional.

## ****20. What Do You Understand by Exterior Angles in Polygons?****

It is the angle between any side of the polygon and the line extended from the next side. It is therefore created by extending one line.  The cumulative value for all the exterior angles in a polygon is 360, handy in related calculations.

## ****21. What Is a Cubic Meter and How Many Hectares Make Up a Square Millimeter?****

A Cubic meter is the unit used to measure the volume of an object, either by using its lengths or diameters. It is written as m raised to power 3. A cubic meter is equal to 1000 litres. A hectare, on the other hand, equals 10,000 squared meters. It is arrived at by multiplying 100 m by 100 meters on each side.

A square millimetre is arrived at by multiplying two millimetre values. A meter equals one thousand millimetres which makes a square millimetre a millionth of a square meter.

## ****22. How Do You Sharpen Your Mathematical Skills?****

I am an accountant, and therefore, part (if not most) of my work involves Mathematics. This has always kept my mathematical prowess and skills at an all-time high. I have also made it a norm to take part in quizzes every day after work. I also take several students through Mathematics over the weekend.

## ****23. What Is Calculus?****

Calculus is a brand of Mathematics that focuses on the finding and properties of both derivatives and integrals of different functions through summing infinitesimal differences. The two main types of Calculus are differential and integral. In simple terms, it is the Mathematical study of continuous change and was formerly referred to as infinitesimal calculus.

## ****24. How Do You Ensure That Your Work Is Accurate?****

I am always keen and attentive to details, which has always helped me maintain high levels of accuracy in all my work. I also go through my work twice to ascertain that I have done everything well and used the correct formulas. At times I also use calculation software to help me arrive at solutions.

## ****25. How Would You Teach an Intern a Mathematical Concept Given the Chance?****

I believe in relating a concept to a real-life situation. It betters understanding and makes the interaction friendly. If real-life situations do not exist, I will explain to the intern in detail and break down the entire concept before tackling each sub concept individually.

Calculus glossary