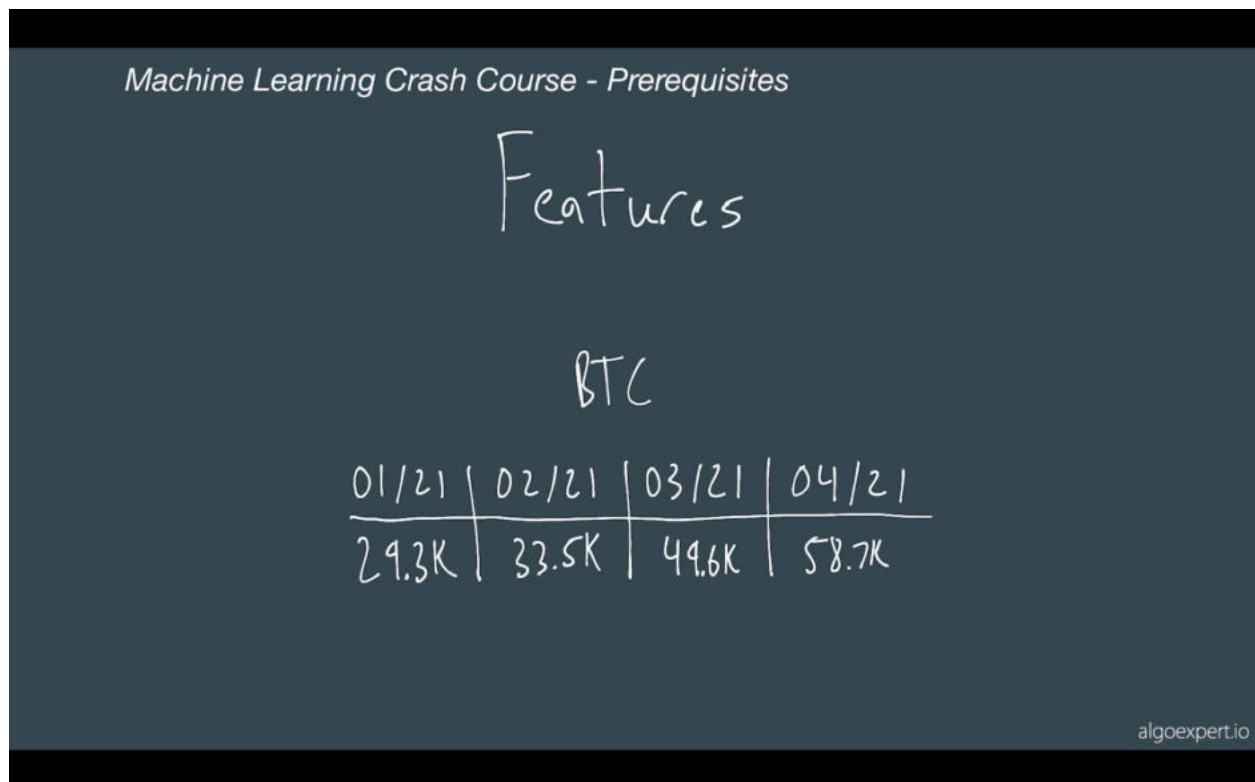


Prerequisites

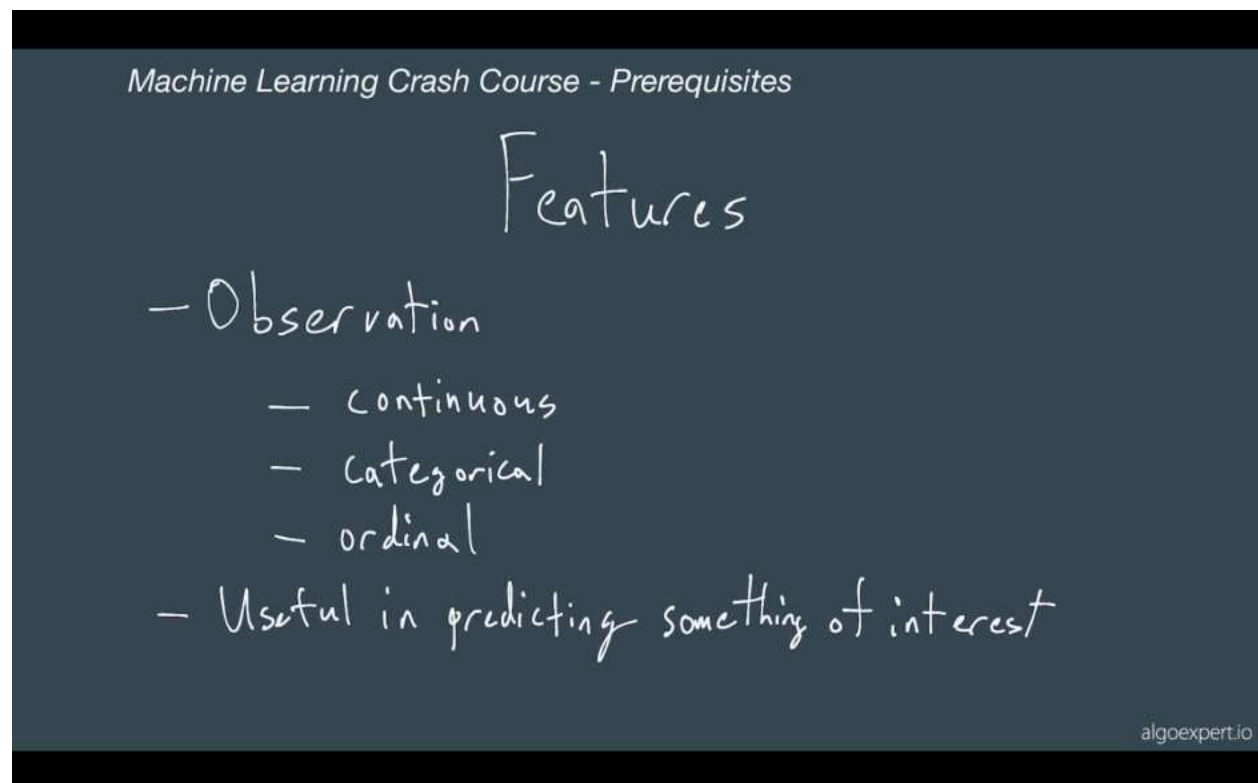
Hello, everybody. Welcome to the machine learning crash course prerequisites video. Here, the idea is to give you some idea of what background could be useful before we begin the machine learning crash course.

Now, if your goal is to just **become conversational in machine learning**, then I wouldn't focus too much on the concepts in this video. However, if your goal is to interview for a technical ML based role then it will be really helpful to know these things to help you grasp the math behind the models.



So first let's go over some terminology generally models we'll take in features. An example of **features** is what we saw in the introduction video which had the price of Bitcoin organized by each month. Features are generally observations that can be **continuous** such as the price of Bitcoin. It could be **categorical** such as the label that we had for Bitcoin which is either up or down, and it can also be **ordinal**. This is the case where it's a category but it can be ordered. This could be

something like small medium and large. Features are useful when we want to predict something of interest.



Features are generally paired with labels. Like we mentioned earlier, the label that we used in the Bitcoin example was a directional indicator of up or down. That would be a categorical label. However, labels can also take on continuous variables as well as ordinal. Labels are generally something that's going to be useful that we want to know. Now, examples are pairs of features and labels. Examples are typically going to be used in a **supervised model**. Unsupervised models will not have the label.

Examples

- Features with a label
- Usually many examples for a supervised model
- Unsupervised models don't have labels

algoexpert.io

Now in terms of math, we'll be going over concepts such as arrays or vectors. In this case, an array is just a feature which has been transformed to a list of values. These could be words as well as numbers. Now, these factors can be assigned to some variable in our case, I'll be using \bar{x} a lot.

Machine Learning Crash Course - Prerequisites

Math

- Arrays / vectors

BTC

01/21	02/21	03/21	04/21
29.3K	33.5K	44.6K	58.7K

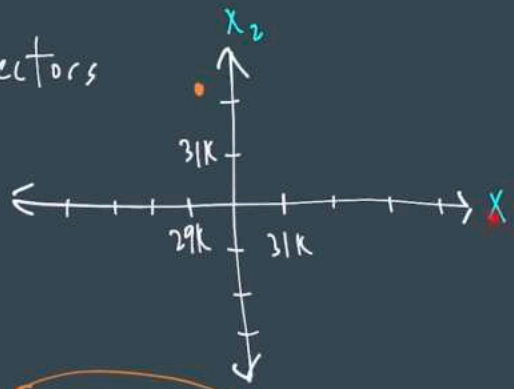


[29.3K, 33.5K, 44.6K, 58.7K]

The bar just indicates that it is in fact a vector and each element in the vector will be called x one, x two, x three, x four, and so on for how many of our values are in the vector. Now, in some cases, what you'll see me do is select a couple of these elements in the vector and actually plot them out. So we could have x one on this axis, x two on this axis. And if we plot 29.3 K and 33.5 K together on these axis we would have a point right here. Again, this is just plotting a vector that we have.

Math

- Arrays / vectors



$$\vec{x} = [29.3k, 33.5k, 44.6k, 58.7k]$$

Now next we can have a two dimensional array and that will give us a matrix. So for instance, this is a matrix and I'll generally assign a variable to some capital letter. So here, this matrix could be A, and that could be matrix B.

Machine Learning Crash Course - Prerequisites

Math

- Arrays / vectors
- Matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 2 & 9 \\ 6 & 5 & 7 \\ 4 & 3 & 1 \end{bmatrix}$$

algoexpert.io

Now, once we have these two matrices what we can do is a matrix multiplication that will look something like this. Now, if we just have a single matrix, we can do the inverse of a matrix that would be notated like this. And then finally, we can transpose a single matrix. Now we don't often use these across the crash course but we do use them a couple of times.

Machine Learning Crash Course - Prerequisites

Math

- Arrays / vectors
- Matrix
 - Multiplication
 - Inverse

$$A^{-1}$$



Machine Learning Crash Course - Prerequisites

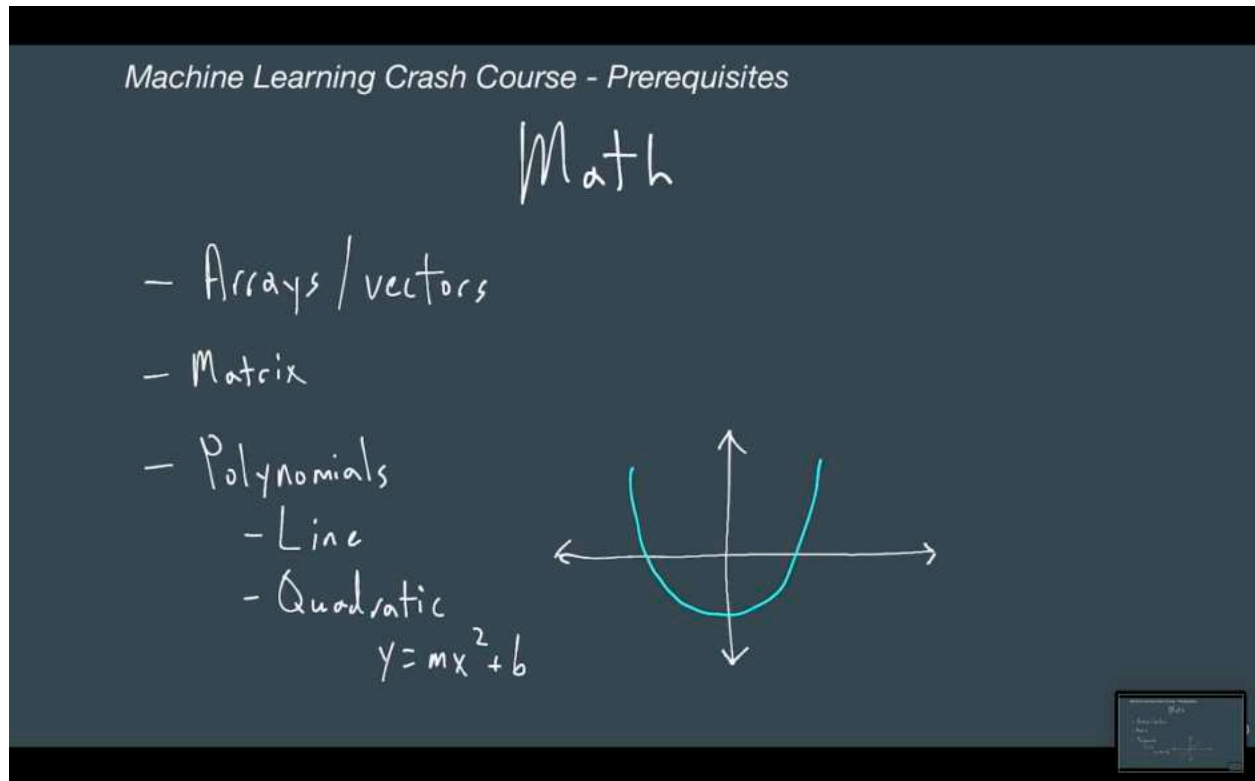
Math

- Arrays / vectors
- Matrix
 - Multiplication
 - Inverse
 - Transpose

$$A^T$$



So I just want to let you know that we will see them, as well, in a couple of videos, we'll be talking about polynomials. Now specifically, we'll be talking about lines. This is just the equation of a line. We'll also be talking about quadratics. So here's an example of that, but in general we'll only be talking about polynomials a couple of times.



Finally, we're going to have to take the derivative of these polynomials for some of the advanced machine learning models, but we'll cover that when we get there. Now beyond the math we also want to know a little bit about probability. So we'll want to know basic probability, for instance rolling a two on a six sided die would have the probability of one out of six, that's because there's one way to get a two out of all of the possible ways that we can also roll the die.

Probability

- Basics

Rolling a 2 with a six-sided die

$$\frac{1}{6}$$

So we could roll a one, a two or three or four and five and a six, which gives us a probability of one over six. Now the probability of rolling two twos with a pair of dice is just that multiplied with itself because both of those have to happen. We will be using the equivalent of an and in the world of probability. So rolling the first two was one out of six and rolling a second two will take on the probability of 1 over 36.

Probability

- Basics

Rolling two 2's with a pair of dice

$$\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Now a little bit more in depth on the probabilities. We'll want to know conditional probabilities. So for instance, the probability of drawing two hearts in a row from a deck of cards would first begin with a 13 out of 52 chance. That's because there are 13 hearts out of all 52 cards. And then once we drew that card out there's only 51 cards now, and 12 of those are hearts that's because we already drew a heart. So here, the probability of drawing two hearts in a row would be 5.88%.

Probability

- Basics
- Conditionals

Drawing two hearts in a row
from a deck

$$\frac{13}{52} \cdot \frac{12}{51} = 5.88\%$$

algoexpert.io

Lastly, in probability, we'll be talking about distributions. So a very common distribution is the gaussian distribution. An example of a Gaussian distribution would be the heights in inches of a population. So let's say the average would be 63 inches. So that would be the most likely height for a person. And then it will get less and less likely for the height to be away from that average. So here, the chance of being 60 inches tall is less likely than being 63 inches tall.

Probability

- Basics
- Conditionals
- Distributions
 - Gaussian



algoexpert.io

Another distribution we'll be talking about is the uniform distribution. When we talked about rolling a dice, every single number on the dice had an equal probability of being landed on. So in that case, it would be a uniform distribution across all of the sides of the dice.

Probability

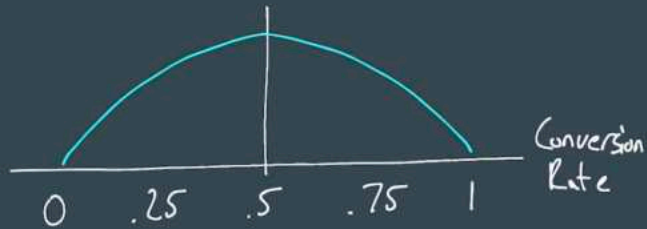
- Basics
- Conditionals
- Distributions
 - Gaussian
 - Uniform



Lastly, we'll also be covering the beta distribution and this is good if we want to model something such as rates. So let's say that we wanted to measure the conversion rate of some model. **Here, conversion rate just means, did they click on the ad? Did they purchase a product? Did they view the video?** Things like that? This will help us understand better. The actual conversion rate that we've measured so far. So in this case, if we've measured that half of our users clicked on a particular ad then the conversion rate would most likely be 50% but there's also a distribution around this 50% such that it could less likely be, 75% as the model continues to operate.

Probability

- Basics
- Conditionals
- Distributions
 - Gaussian
 - Uniform
 - Beta



algoexpert.io

It could also be less likely to be 25% conversion rate as the model continues to operate. And we get down to extremely unlikely for the conversion rate to be zero or 100% based on the model's performance so far. We'll talk more about beta distributions once we get there that's all that we'll need to know for now. Let's go ahead and begin.

Key Terms

Features

A set of quantities or properties describing an observation. They can be binary like "day" and "night"; categorical like "morning", "afternoon", and "evening"; continuous like 3.141; or ordinal like "threatened", "endangered", "extinct", where the categories can be ordered.

Labels

Usually paired with a set of features for use in supervised learning. Can be discrete or continuous.

Examples

Pairs of features and labels.

Dimensions

Here, the number of features associated with a particular example.

Vector π

Here, a feature vector, which is a list of features representing a particular example.

Matrix π

An array of values, usually consisting of multiple rows and columns.

Matrix Transpose π

An operator that flips a matrix over its diagonal.

Polynomial π

A function with more than one variable/coefficient pair.



Derivative π

Indicates how much the output of a function will change with respect to a change in its input.

Probability π

How likely something is to occur. This can be independent, such as the roll of the dice, or conditional, such as drawing two cards subsequently out of a deck without replacement.

Probability Distribution π

A function that takes in an outcome and outputs the probability of that particular outcome occurring.

Gaussian Distribution π

A very common type of probability distribution which fits many real world observations; also called a normal distribution.

Uniform Distribution π

A probability distribution in which each outcome is equally likely; for example, rolling a normal six-sided die.