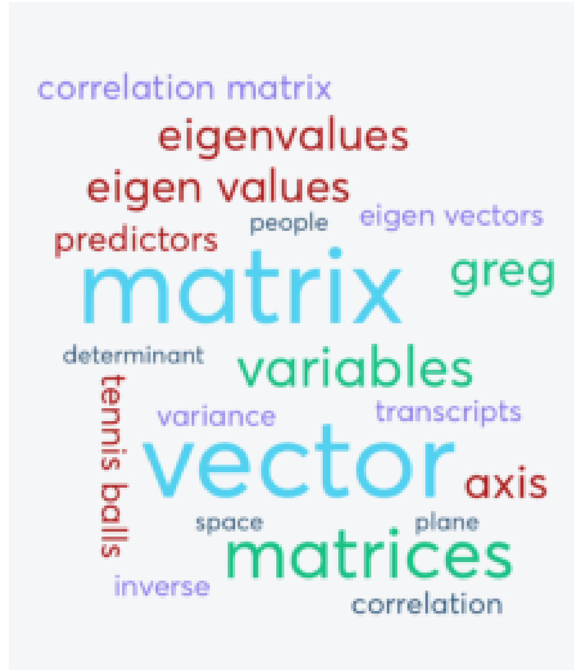
The Podcast *Quantitude*

with Greg Hancock & Patrick Curran

Season 3, Episode 23:

The Mättrix Part 2:
Using Matrices to our Advantage

Published Tuesday March 8th, 2022 • 53:57



SUMMARY KEYWORDS
matrix, vector, determinant, matrix algebra, inverse, column, people, multiply, called, scalar, variables, talking, dimensions, data, represents, elements, third eye, diagonal, point

Patrick 00:01
Welcome, my name is Patrick Curran and along with my geometrically multi dimensional friend Greg Hancock, we make up quantitative food. We are a podcast dedicated to all things quantitative ranging from the relevant to the completely irrelevant. In this week's episode, Greg and I continue our discussion from last week in our sequel The Matrix part did, in which we continue to explore the magic of matrices including estimation and Eigen values and Eigen vectors. Along the way, we also discuss flawed audio transcripts 50 Shades of Greg drunkenly shoving a matrix drug mules things you need isomorphic interdigitation plywood and tennis balls, heroin Filton, condoms, talking to volley balls, bada bada bang Dang. Diggy diggy meat grinders not going to prom vector bouquets and the right stuff. We hope you enjoy this week's episode. So you do know when it comes to movies, as a rule, sequels suck. There are some exceptions, but yeah, for the most. So let's see. Jaws to suck rocky to suck Back to the Future to kind of like that one. Just got to come back with me. To the future. Yeah. Okay. So we'll put that in the acceptable the exception of the rule. There are always exceptions to the rule is the Godfather Part Two Oscar winning. That was Oscar winning and arguably better than the first one.

Greg 01:39
In the good pile, Dark Knight from the Batman reboot series.

Patrick 01:42
I don't like any of the Batman. So yeah, sure.

Greg 01:45
What whatever. How about Star Trek to Wrath of Khan? I

Patrick 01:49
don't like the Star Trek movies either.

Greg 01:52
What about the we get this right? The matrix to

Patrick 01:57
the matrix? Yes. So why are we talking about this? Well, we are doing a sequel here. To avoid getting sued, we made some reference to a very famous prior movie starring Keanu Reeves about taking a red pill or a blue pill and going into a structure that we refer to as the matrix.

MovieClip 02:19
This is your last chance. After this, there is no turning back

Patrick 02:23
just to avoid getting sued. Now, there was a matrix part two. The problem is is you and I did not plan on doing a sequel, we ran out of time and I had to go to the bathroom.

Greg 02:37
You're like rocking back and forth.

Patrick 02:41
Like, we still have Eigen values and Eigen vectors and spectral decomposition to go. And so we call it right is sometimes you just have to call it we have a sequel, this was somewhat unplanned, but I got an idea. All right. Usually, when they do the sequel, you have to set it up, where you bring the people on board who maybe didn't see the first one. So you've got some overview of what happened in the story. Now, this is a serious thing. And I'm actually very excited about this. We are now posting transcripts of the audio on the webpage. Alright, so this is very cool. They are imperfect, they are lightly edited. I do not sit down and try to correct all the weirdo things that you say that I say, here's what I thought we would do is we weirdly have not edited the MOT tricks, part one. That's right. So we're in that weird time machine where audience you now are listening to this as it's posted after the matrix Part one was, but as Greg and I are recording this, the raw audio is actually sitting on a hard drive. And by the way, Greg, you probably want to get on that because he knows that in a matter of days, but I'm not gonna micromanage because

Greg 03:57
I have 20 hours just lying around in my schedule. Exactly. It tends to reside

Patrick 04:01
between 11pm and 6am increments. I had an idea, okay, let's take the raw audio, run it through the transcription, and there's a voice recognition component to it. That speaks it. Oh, nice. We're going to have the audio read the unedited transcripts, and we're going to use that as the summary of part one. As we move into part two. I like it. Okay, so here's the original clip before editing.

Greg 04:30
Okay, so that was confusing. Let's have a new start where we refer to it as a scalar.

04:36
Okay, so that was confusing. Let's have anus chart where we will refer to to scaly hair.

Greg 04:44
Okay. That's before like light editing.

Patrick 04:49
Okay, that was the raw audio that hasn't been edited for posting. Added did AI transcript. Okay, okay. Yeah. No, I'm not sure how helpful that was. Bah, we can have an anus turret in a river today. If we would like. Should we try another one?

Greg 05:11
Yeah, there might be a better one.

Patrick 05:12
We're gonna try that again. Maybe that was just a blur. All right. So we'll try this one here. We'll hit play again. So Greg, you've told me before that you'd like it done in a matrix so I'll lie there.

05:25
So Greg, you've told me before that you like a dominatrix? I'll let her

Greg 05:34
I didn't know you're gonna go all 50 Shades of Greg on me hear

Patrick 05:38
me is that actually is a correct transcription?

Greg 05:43
Was that our episodes? That's another conversation.

Patrick 05:46
That was maybe a hot mic moment. Okay, I'm starting to question the technology,

Greg 05:53
quietude the podcast that requires a safe word,

Patrick 05:57
or safe word is a new start.

Greg 06:00
Maybe this is a pattern detecting and the ability of the software to recognize.

Patrick 06:06
Okay, this is a little too early for primetime.

Greg 06:10
So I hope people out there find those.

Patrick 06:13
No decoding clarify, this is not the quality of the transcript that's online. This is the reading recognition. So we really do have useful transcripts that don't involve dominatrix or anus tarts,

Greg 06:26
although people might prefer both versions. How about this,

Patrick 06:29
how about if I do it, cliffnotes of what we talked about, in just terminology, I'm going to do basic definitional, and you come back with the geometry of it. That sounds good. Alright, here we go, we are going to start with what's called a scalar. A scalar is just a number in part one, Greg had a wonderful description that is called a scalar. Because it rescales things, a collection of scalars in a column, so two or more, we're going to call a column vector, we're going to denote that lowercase bold, we're going to define the postal code of that as the rows by the column and so it's n by one would be a column vector of length. And if we want to move to what's called a doubly ordered organization of vectors, we move to a matrix. So a matrix has multiple rows and multiple columns, we're going to use uppercase bold, so instead of a lowercase x, we would have an uppercase x, and this now has rows of n and columns of P would be an n by p matrix. When we want the postal code for an individual element, we go the row first and the column second. So the three four element is the scalar. That's in the cubby hole in the third row. And the fourth column, a matrix can be rectangular, a matrix can be squared, we can drunkenly walk up to the matrix or the vector for that matter and shove it on its side. So if it's a matrix, the first column becomes the first row. The second column becomes the second row, we call that A transpose transpose is a mathematical term for drunkenly shoving a matrix. We can do that to a vector as well, and a column vector flops over on its side and it becomes a row vector, we can compute what's called a determinant. It's a single numerical value, and it represents the generalized variance in a matrix. If it is zero, that means there's some linear dependency within the matrix. Where do we often do these is we will do these with what are called symmetric matrices that is below the diagonal values are the same as above the diagonals, whatever whack ability world would we ever do that in? Anytime we work with a covariance matrix, or with a correlation matrix? Those are symmetric. What else can we do with a determinant? Well, we can compute what's called an inverse of a matrix. Why would we do that? Well, an inverted matrix has the same dimensions as the target matrix. So if we have matrix A, and it's p by p, then we can invert a and that's also P by P. Well, turns out in matrix algebra, we can add two matrices in the usual way, we can subtract two matrices in the usual way, we can multiply two matrices with a little goofiness that we described in part one, but it's pretty straightforward. There's no straightforward corollary to dividing two matrices. So we're going to trick it instead of dividing matrix a by matrix B, we're going to multiply matrix A by the inverse of matrix B. And that's going to do the same thing as dividing. I think those are the main features

Greg 09:45
Alright, so I'm supposed to do the geometric version of what you just did. What I would like to do is hit on some key points from last time once it will be especially useful here in the matrix part do

Patrick 09:59
or what Whatever Metrix Yep,

Greg 10:03
here's some of the key ideas that I really want to draw upon. And it's going to start with the idea of your data matrix, you have a data matrix an n by P data matrix P variables going across the top as columns, and N cases going down as the rows, each one of those columns is information about a particular variable. One way to think about that geometrically, is imagine that we have an axis system, a coordinate system, where we have one axis actually, for every person. And as I mentioned, last time, we might have an axis for the person named Al and an axis for the person named Beth and an axis for the person named Carl all the way through all the people that we have. Now what I do is I am going to put a point in space that corresponds to that first variable, and I draw a line from the origin 00 point of that space out there, a little arrow that goes there, that's a vector, that vector represents the information on that first variable, but I can do that for the second variable, and the third variable and the fourth variable, if the second variables arrow, the second variables vector is in the exact same place as the other one, then those two variables are perfectly correlated, the second variable doesn't really contribute any different information from the first one, if that variable is at a right angle to that one. And what I mean by that is that if the vector formed by variable 1x, one and variable 2x, two are at right angles, it means they communicate completely different information from each other, they would have a 90 degree angle between their vectors. So one of the key ideas from last time is that we can think about variables as being these things that we call vectors in a space that's defined by the end dimensions of the people or the cases that we have. And the angle between any two vectors tells us something about how redundant the information is, and what we call that as the correlation, right. And quite literally, the angle between two variable vectors, if we take the cosine of that, that is the Pearson product moment correlation coefficient, Mike drop, it is just so beautiful. If I say that a little bit differently, if we took one variable, and we projected it on to the other vector, so that it made a right triangle, the length of that projection on the other variable would represent the correlation. So the idea of variables having angles and those angles capture how similar the information is in those variables. And the idea of projections are really foundational matrices are really just things that move those vectors around, when we multiply a vector by a matrix, it transforms it in some way, it might move it to a different location in space, it might stretch it, it might shrink it. So one useful way to think about matrices are as these transformers that move our systems of vectors around. And I think that's about all I want to say geometrically for now. Now, we're

Patrick 12:51
going to go into new territory from the part one, because there's so many interesting things that we didn't talk about, there are a couple of things that I am particularly intrigued about, we can use these matrices to compute all sorts of stuff for us, right. So in a way a matrix is like a drug meal. So we want to get something done, we want to get it done efficiently. We don't want to do it ourselves.

13:15
Wow, the drug company won't do anything to help us all the drugs we want. Right across the border, I have a friend who will help us

Patrick 13:24
there's a drug mule aspect to this. Picture a vector of outcomes, right, let's just say we have a sample of 10 called led why. So it's lowercase, it's bold. It's a vector of 10 scores. Now we're going to create a vector of just ones, we can multiply those two vectors, and it's going to be one times y one plus one times y two plus one times y three, we can use that vector multiplication to get the sum of our 10 y scores. Yeah, instead of just ones, we can have the inverse of n. Well, now it's the inverse of n times y one plus the inverse of n times y two, we computed the mean, well, what if you have more than one variable? Okay, well make y a matrix, make the vector of ones now a matrix of ones do the same thing. And now we have a vector of sums, we have a vector of means. And we can just keep doing this we can calculate means we can calculate variances, we can calculate sums of squares and cross products and covariance matrices and correlation matrices. And they're all just based on the data matrix as Greg just described, right? Everything we have, think about what you have in the world, and we have an n by P data matrix. That's it our entire lives, our happiness, our tenure or contributions to society is our worth as human beings. Frankly, our reason for living is an m by P data matrix. You do growth mixture models you do discriminant function and out assists you do fill in whatever you want. It's all an MIP data matrix.

Greg 15:04
So those matrices and vectors are really like these little workhorses that do things for us. And one of the things that they do for us starts right with some of the ordinary least squares regression, some of the general linear model stuff that we do. Right,

Patrick 15:16
exactly. Let's use that as just one example of a very large number of examples where we don't appreciate matrices as we should, there are a couple of themes to this episode one is matrices or drugged meals. And the second is, we do not show our appreciation for matrices enough. Let's start big, you've got one predictor and one outcome in a regression. So 1x and one y, what information do we have available to us we're not going to talk about means I'll just focus on variances and covariances means scale in as they would in the usual way. We have a variance of x, we have a variance of y, and we have a covariance between X and Y. That's all you have in your entire world. What is it from the Steve Martin movie, and this

Greg 16:02
ashtray? And this lamp?

16:05
Yes, train a power game and the remote control.

Patrick 16:11
So all I need is this variance of x, this variance of y, and this covariance of X and Y. Alright, that's all we have in the world. A variance of X variance or Y and a covariance between X and Y somebody someday in the past taught you that we're going to regress y on x, right? We always do it backwards. It's the dependent variable on the independent variable, meaning x predicts y. What is that regression coefficient? That regression coefficient is equal to the covariance between X and Y divided by the variance of x, that gives us our OLS regression coefficient, the covariance between X and Y divided by the variance of x. So we're simply rescaling the covariance in a rise over run interpretation that we all want. Alright, well, how do we use our drug meal? To help us scale that up? Well, let's take our predictors. Now, in action, we have multiple predictors. Let's say that we were to have P predictors. So now our data matrix is m by P, that's X, how many predictors you might ask, doesn't matter. The beauty of matrices is it's n by p. If you have to, well, there are two columns, if you have 22. Oh, wait for it. They're 22 columns, they're just balloons, they expand to the number of cases you have, and the number of variables that you have. That's our data matrix x. Now, in a multiple regression, we still only have one dependent variable y. So little y is just going to stay a vector, because whatever it is, we're predicting we only have one. So what we have to do is do a matrix expression to do the equivalent of the covariance between X and Y divided by the variance of x. Now, if you want to isolate yourself at a party, as Greg and I've done repeatedly, this is an equation that is handy to have at your fingertips, because it will ensure that you don't have to talk to anybody, not a soul about a hollow and vapid conversation that you don't want to have with a person you don't want to talk to. Ready everybody, x prime x inverse x prime y. Well, what the heck did we just do? Alright, repeat after me for driving x prime x inverse x prime,

Greg 18:31
why not one of the better band names honestly,

Patrick 18:34
it's not isomorphic interdigitation. Greg and I had a wonderful opportunity to meet with a group yesterday, it was headed by Aidan right at Pitt, the term isomorphic interdigitation came up that's a band name. Axis your data matrix and by P let's start with that first part, x prime, you drunkenly shove it on its side. It's n by P. Well, now it's p by n. Well, what does x prime x do? Well, it's a p by n times m by p matrix. So we get a p by p matrix, x prime x gives us our sums of squares and cross products. That's the same as a covariance matrix, we just haven't divided by n minus one. Well, what is x prime why we drunkenly shove X on its side, multiply it by y. Well, that's the sum squares and cross products, again, insert covariance matrix between our predictors and our outcome. So x prime x is our covariance matrix among our predictor x prime. Why is our covariance matrix between our predictors in our outcomes x prime x inverse? Well, we talked just a moment ago. That's our tricky way of dividing. So we're dividing by the sum squares in cross products matrix of our predictors. What are we dividing? Well, x prime y, which is our sum squares and cross products of our predictors in our outcomes? What is it for one predictor? It's the covariance between X and Y divided by the variance of x. What is it when we scale up to matrix is the covariance between X and Y divided by the variance of x, but it's just scaled up to as many predictors as you want. Wow, what is the matrix expression for two predictors in 100? Cases, x prime x inverse x prime y, what is it when you have a million cases and 500 predictors? Gregg,

Greg 20:30
x prime x inverse x prime y, beautiful,

Patrick 20:33
I'm guessing you can redo what I just said. But using geometry,

Greg 20:39
oh, boy, can i

Patrick 20:42
You're way too excited for that. I'm completely

Greg 20:45
geeked. About this, here's a way to think about it. And what I would like to do is just go straight to the multiple regression case. And think about this from a geometric perspective. For us to be able to visualize it as best as possible, I'm just going to do two predictors, we've got x one and x two, and then we have a Y variable. One way, we often think about this, in addition to the very cool matrix ways that Patrick told you, we are looking to make some linear cocktail out of the x's to try to get it as close to y as possible. And that linear cocktail out of the x's we call y prime or y hat, the predicted Y. And we talked about it being as close to y as possible in terms of having minimum residual sums of squares between the two, or maximizing the correlation between y hat and the actual y. So there are different ways to think about it, you have x one and x two and y. And I want us to think about those as vectors in space, just like I talked about. So we have this n dimensional space with each dimension corresponding to a person or a case and we plot x one in that space. And it is a vector that zooms out to the point that defines everybody's score, we do the same thing for x two, it is another vector, maybe those vectors are really close to each other, because x one and x two correlate really highly, maybe those vectors are almost at right angles, because x one and x two contribute very different information. But they are out there in this n dimensional space defined by the cases that we have. Well now enter y, y is just another vector in the space y doesn't care, right, we could have called it x three, it doesn't matter. It's just some other vector in that space. The distinction between the x's and the y is what we intend to do with it. And what we intend to do with it is to try to take the information in the axes and get as close to the Y as possible. Or we could say that the other way, take the information in y and get it as close to the x's as possible. So what I want to do now is I want us to think about this subspace, that sounds kind of tricky, but not really, I want us to think about the subspace defined by x one and x 2x. One and x two are just these vectors in space. And we could draw a plane through them. Unless they're exactly right on top of each other, in which case, we've got a colinearity problem that's going to thwart anything that we want to do. So let's assume that x one and x two are not right smack on top of each other, and we put a plane through them. Now what that plane actually represents, you can think about that as all the possible vectors from the origin that are in the same plane as x one and x two. Another way to say that is that all of those vectors are all the possible linear combinations of x one and x two, we could combine x one and x two in any kind of linear cocktail that we want, Wait x one, Wait x two, and we could create all the other vectors in that plane. So a good way to think about that plane is all possible combinations of x one and x two, now, there is y, and y almost certainly is not on that plane, if y was on the plane that x one and x two make, then y would be perfectly multi collinear with them. And we would be able to use

x one and x two to perfectly capture y. But y is almost certainly not on the plane defined by x one and x two. So the question is how close is y to the plane that's defined by x one and x two. And if you think about just maybe the desk that you're sitting at or the dashboard of the car that you're driving in right now, or whatever, think about that as the plane that is defined by the x one in the x two vector, and y is some vector that's just poking up out of that plane. Well, where is y closest to the plane? And the answer is directly below y. If you dropped a little line directly down from y, you would get Y's shadow in the plane that is formed by x one and x two, that shadow of Y is Y hat. That is where y is as close to the information in x one and x two which is defined by their plane as possible. That vector that is the shadow of Y is Y hat and that Y hat is some combination of x one and x two. So In all the matrix stuff that Patrick is sharing with you, that's exactly what's going on with the x prime x inverse x prime y is figuring out what is that cocktail of x one and x two, that defines the shadow of y in the plane defined by the x's. And there's some cool features about this. If we measure the angle that that variable Y has to its shadow y hat, that angle, if we take the cosine of that is the multiple correlation. It's a Pearson correlation between those two things. But it's also a multiple correlation as we think about it in terms of multiple regression. So the closer y is to that table top that's formed by x one and x two, the higher the multiple correlation, the big R, the farther away it is, the more that Y variable vector is orthogonal or perpendicular to the plane that's defined by x one and x to the tinier at Chateau. And that means that x 1x, two really aren't going to be able to inform you a whole lot about y, because y barely relates to those at all, the description that I just gave you is one that you can visualize in three dimensions, because two of the dimensions are x one and x two, and the third dimension is y. But you could have any number of axes write P x's, that's harder to visualize. But you can still imagine that there is now not a plane that is spanned by this collection of axes, but there's more of a space that is spanned by them. And when we ask about the relation that the x's have to that loan, why what we're really asking is, why is out there in some direction relative to these x's were in that space that is spanned by all of these x vectors, is why the closest what we do in multiple regression, when we have more than two predictors is we're trying to find that location in the space that is spanned by the x's that is as close to the Y vector as possible. And the angle that that Y vector forms with that closest place in the space spanned by those x vectors. That angle is once we take the cosine, the multiple correlation, and the projection of that y onto the now space that's defined by those x's, that is y hat, that is the best we can do the best y looking thing that we can make out of the x's that is in the space that they span. So that's the geometry of multiple regression. And it's exactly exactly the same stuff that Patrick is talking about with all of the matrices, although you left

Patrick 27:32
out part of the matrix manipulation involving tennis balls and plywood. So one way I tried to describe it is everybody picture in your mind's eye, you're in a room and you go to the corner, and on the floor coming out against one wall is your x one axis, going out on the floor, on the other wall is the x two axis and going up toward the ceiling is the y axis. Now if you have two predictors, everybody has three values, their value on x 1x, two and on y. So each subject holds a tennis ball, and you start at the origin, and you walk along the first wall on x one, two, your value, then you walk out into the floor, right as you're parallel to the x two axes, and you stop at your x to value. And then you hold the tennis ball up to where it corresponds to your y value. And we're going to hang the tennis ball there, the whole sample is going to do that. And we're going to have this chandelier of tennis balls that represents that x 1x, two y space, then a couple of people are going to pick up a sheet of plywood, and they're going to

try to run it through that swarm of dangling tennis balls to be the air quote, best fit where you're going to nail it on the X one axes, you're going to nail it on the x two axes. And that's going to cut through that tennis ball swarm. And what we're going to do is what Greg just described as minimize the sum of the squared residuals. I'd like your shadow analogy, Greg, I've not used that of where, if you have a light above or below is where does that shadow of the tennis ball project onto that plane. Then when everything is nailed in place, and you step back, where you have three pieces of information, you have the slope of the edge of the plywood on x one, you have the slope of the edge of the plywood on x two, and where those two corners of the plywood nail when to the Y is your intercept, you have an intercept in two slopes,

Greg 29:45
it all comes down to dangling balls with you. Each ball has a distance above or below that plane. And so those are the individual residuals that's just such a great way to think about it's another geometric representation that ones in terms of the point clouds All right that Patrick is helping you to visualize very practically, and I talked about it in terms of the vector space are two ways to visualize multiple regression that go together nicely.

Patrick 30:08
Now, since we're doubling down on visualizing things that are almost impossible to visualize, I think we should complicate things. Oh, why not? Everything we've done so far has been really straightforward matrix algebra, you add things, you subtract things, you multiply things, one thing that we realize is wow, we should be a lot nicer to matrices than we are x prime x inverse x prime Y, and you've got your vector of OLS. slopes for any number of cases. Any predictor, any situation, if you're abandoned on a deserted island, you want x prime x inverse x prime. Why? Because that's all you need. That's job one. Right, I'm not going to talk to volleyball, I'm going to talk to the universe of a product,

MovieClip 30:57
I would rather take my chance out there in the ocean than to stay here and die spending the rest of my life talking.

Patrick 31:07
So let's turn to something that is simultaneously wickedly cool, and mind numbing ly complex. And that is a eigenvalues and eigenvectors who even I with plywood and tennis balls and drug mules. It's hard to come up with a muppet like a description of eigenvalues and eigenvectors,

MovieClip 31:35
Dr. Bunsen Honeydew, here at Muppet Labs where the future is being made today.

Patrick 31:40
So we're gonna have to kind of double down a little bit and push through this. However, when we pop out on the other side, it is insanely cool. And is a tool that all of us use almost every day in some way or another in this kind of work. For sure. We're gonna start out with what's called diagonalization of a square matrix. Already, your eyes are glazing over. I'm not talking to you, audience. I'm talking about Greg. I'm looking at him and his eyes.

Greg 32:07
Here present,

Patrick 32:08
what does that mean? All right, imagine that we have a square symmetric matrix, let's just make it a correlation matrix. So let's call it r. So you have p variables, that's p by p to say that we're going to diagonalize. That means that we're going to express that correlation matrix R as a matrix product. And it is going to be V d V inverse. Well, what does that mean? If we say r equals V d V inverse, we already know from our conversation, oh, we got some multiplication going on, we've got division going on. What we're doing is we're taking our correlation matrix into the garage running the door down, we're going to turn to Kid Rock this time, because we played Green Day last time, oh, I told you the story where my kids were very young, and we have Pandora and I was cooking dinner. And all the sudden Bob did a bada bang, bang diggy diggy came on in my kid thought that it was Kid Rock like rock for kids. And it was actually my lifting music. And I threw myself across the room tearing wires out of the stereo. So we're gonna have Kid Rock bonded to bar.

Patrick 33:24
We're going to reexpress our in terms of two new matrices, V, and D. Well, why on earth? Would we do that? Well, Greg's gonna tell us why we're gonna do that in a few minutes. What I'm going to very briefly describe is what's happening while Kid Rock is playing. And we're breaking down our It is literally a re expression. All we're doing is taking a correlation matrix R. And we're re expressing it in terms of matrix V, and matrix D. What is matrix V? We're going to call those Eigen vectors. It's a square matrix. So what is D? It is a diagonal matrix, meaning that it has values on the diagonal, but zeros everywhere else. And that contains what we're going to call Eigen values. So when you think about Eigen vectors and Eigen values, those are two components that allow us to break down a correlation matrix or a whole variety of other matrices for that matter, but in our example, it's correlation. We're gonna break that down into these Eigen vectors and Eigen values, the calculation is complicated and involves determinant set to zero to avoid trivial solutions, there are two roots to these problems. And in my multivariate class, I would have people calculate them for a two by two matrix, because beyond that, you really need to turn to the computer, but what you get when you do this diagonalization is we have traded in one matrix for two matrices, Eigen vectors and Eigen values, what are those Eigen value? values represent, we can think about it as the Eigen value is the first component in a matrix of maximal variance, I'm going to ambush Greg and have him talk about this a bit more in a moment. It's a composite of our variables that maximizes the variance in that first weighting. And those Eigen vectors are actually the weights that give us that first composite, then the second eigenvalue is the second composite with the vector weights the third the fourth, and we can extract as many Eigen values as there are variables. So, if we have a P by P correlation matrix, we can compute P Eigen values, and each Eigen value has an Eigen vector that goes with it. Well, these Eigen values have some crazy cool properties first, the sum of the eigenvalues that's equal to the sum of the diagonal elements of your target matrix. So for our last week, Greg described a trace and it's the sum of the diagonal, you take the trace of the eigenvalue matrix that equals the trace of your target matrix. All we've done is taken that total variability on the diagonal, and we've just rearranged it, it's exactly the same, right? You can't create matter, you can't destroy matter. Well, we can't create variants. We can't destroy variants. But

we can reorganize the product of the eigenvalue, you ready for this? That's the determinant of the matrix. I know, right? Yeah. Now you can start to see a shimmer of ooh, I wonder how I could make use of that. Because remember, we said, well, when the determinant equals zero, that there's a linear dependency among the columns and rows of our matrix? Well, what's the only way that we can get a determinant of zero? If we think about it as the product of the eigenvalues, what one of the eigenvalues has to be zero? Well, what does that mean? Well, it means that there's no unique variability left when we get that to create by weighting or variables. When we get to that eigenvalue, it's the same thing. There's a linear dependency, if there's nothing left, you're out of gas, you get to that last Eigen value, the Eigen value is zero. That means there's no unique variability without weighting of the variables. You take the product, your determinant, pack them up, smoke them if you got them, but you gotta go home. Now, some terms that you might have seen is the number of nonzero Eigen values that you have, we call that the rank of the matrix. So maybe in your reading, you've seen that a matrix is assumed to be full rank, what that means is there are no zero Eigen values, a matrix is full rank, if all the Eigen values are greater than zero. How cool is that? It's just beautiful. Okay, so we see a correlation matrix, we really like we hotwire it, we drive it into the garage, run it down, sand it, paint it, bring it back out, and we no longer have our but we have matrix V of vectors, and d of values. So what what is that repackaging do for us?

Greg 38:15
I'm going to answer your question, but not before dragging people through the geometric muck. How about that?

Patrick 38:22
Do you need to borrow my tennis balls in my plywood? Because I've got him here? I also have some balloons and inexplicably a condom filled with heroin. I'm good. Good. Okay, just let me know if I can be yours.

Greg 38:38
Yeah, no, I always appreciate that, I want to go back to the space that we have that has variable vectors in it. So it is that n dimensional space where the axes are defined by the cases are people. And each variable that we have is a vector in that space. Imagine many vectors, right P vectors. But I want you to think about them all heading in really, really similar directions, right? So they form almost this, I want you to try to visualize it, almost like this bouquet, or this cone like shape, where they're all heading out very, very similar directions. So if someone asked you to describe where those are going, you would sort of point in the same direction, say, well, they're all kind of heading up there. And if you had to create a vector to stand in representation of that beautiful bouquet of vectors all heading sort of in the same direction, what you could do is you could run a vector right down the middle of that bouquet, it wouldn't necessarily be on top of any one of those, but it would be sort of as close to all of them as you could possibly get. That particular axis actually represents the first principal component, the axis that is in the space spanned by all of the variable vectors, that is as close to all of them as possible that is trying to stand in representation of all of them. But that doesn't capture all the information in those variables. I could put another axis in there that is orthogonal to that one, right? That is at a right angle to that one, just as that first axis captures as much of the shared variability that those original variables have sort of in that direction, I would put in a second axis that captures the variability that those have in

some other dimension. And then I could put in another one, that would be in a third dimension, fourth dimension. And I could do it all the way until I have exhausted the p dimensional variability in these p variables. So what have I done, I have taken the variability that this has, we'll assume, for right now in p dimensions, and I have repackaged it into a space that now is made up of vectors that are completely orthogonal to each other, the original x vectors, not orthogonal to each other in all likelihood, but these vectors are orthogonal to each other, oh, my gosh, this is just going to get so cool, the length of each of those vectors is going to be tied to the eigenvalues, the length of that first vector that splits that beautiful bouquet of vectors, it's going to be really long, because that captures for the most part, the direction that those all are heading, that vector will have the greatest length, technically, the length of that vector is the square root of the Eigen value. The second vector has a link that is the square root of the second eigenvalue, the third vector, the square root of the third Eigen value, and so forth all the way. If someone said, what is the volume of that big rectangle? It wouldn't be a rectangle anymore, right? If it's in p dimensions, it would be the P dimensional equivalent of a rectangle, what would be the volume of that rectangle? And the answer would be well, I would take the length of this side of this three dimensional rectangle A times this side times this times this side, I would take the product of the square root of all of those eigenvalues, and I would get the volume of that, well, you know what the idea of this space is that it contains the same amount of information as the original variable vectors. So when Patrick was talking about and this is a, it's just all comes to, I'm getting Misty just thinking about it. So when Patrick talks about the determinant of the correlation matrix representing some measure of generalized variance, we can think about that as being some measure of volume, how dispersed all of the information is. And the smaller that is, the smaller the amount of variance that has, what that translates to into this vector space is that the more all of the variable vectors are heading off into the same direction, the less variability they have in multiple dimensions, the smaller the determinant gets. And what we talked about last time is that if we think about the parallelogram that those vectors would make, or multiple dimensions is called a parallel a pipette, the more those vectors are aligned, the smaller the volume of the shape that they would make, the smaller the determinant that they would have, well guess what the volume of that would be the square root of the determinant of the original matrix, and it would be identical to the volume of this new rectangular shape that we're making with these new axes that are the first second, third, fourth, etc principal components. So, all we have done is repurpose this space, we put in a major axis to start the captures as much information as possible, we put in a secondary axis that captures as much information as possible in the next dimension while being orthogonal to the first etc, we just took some funky shape and rectangular arised it that is all the Eigen value Eigen vector decomposition is doing the Eigen vector tells you where the axis is, and the Eigen value tells you the scale of that axis, how long or how short that axis is, in a correlation matrix. When you have redundancy among the variables. By the time you get down to those last few Eigen values, they could even be zero. And what that means is that there is multicore linearity among these variables that even if there are p dimensions worth of data, those things are lining up in multivariate space to some extent. So, if you had only three variables x 1x, two and x three, but they all fell into a plane, your three dimensions worth of variables would effectively be only two dimensions worth of information. In that case, your third Eigen value would be zero, because after you fit that plane with a first axis and a second axis, there's no need for a third axis, or you can think about there being a third axis, but it has length of zero. So all of this ties together so beautifully the Eigen vectors, redefining the space of your variables, the Eigen values telling you something about the length of those and those map perfectly on two principal components.

Patrick 44:35
I've two observations, one you gave me static about balloons, tennis balls, and plywood, yet it's a beautiful bouquet. I'm just saying Far be it for me to point out hypocrisy, but Okay, uh huh. We do have a beautiful bouquet and listeners you can't see of course, but as he was describing it, he had a his fingers and hands in this beautiful bouquet opening up. So you're going to have to imagine that as well of him holding his prom flowers. I didn't go to prom. The other one, though, is as soon as we start thinking about matrices and thinking about things in the way that Greg just described, some things, at least to me start making a little more sense. And one example of many is multicollinearity. A lot of times people will say, Well, I don't understand. I'm having trouble inverting the matrix. But I don't have any correlation that's greater than point six. In my typical responses, I don't care, mostly because that's how I show support, right? That's a Bivariate Correlation, the very term Muay Thai colinearity. Greg is waiting for his date to pick him up on prom night with his beautiful bouquet of flowers. It's not one correlation. That's right, right is that this whole notion is you have this explosion of vectors that are going out. And that the multi colinearity part is by the time you do what Greg described, you take the first component, and then you take the second one that's orthogonal to the first thing, you take the third one that's orthogonal to the second, and you get to the last one, and there is nothing left. That's multicolor linearity.

Greg 46:29
And that highlights one of the key purposes of what we do in so many aspects of data analysis. That's that notion of data reduction that even if we have p variables, that doesn't necessarily mean we have p dimensions worth of information. And even if we do have p dimensions of information, we might not care about all p dimensions worth of information. And when we talk about principal components as a data reduction technique, or as Doug talked about, right, this is one big sausage maker, I think that's that might be or he called it a meat grinder, right?

Doug Steinley 47:02
I always think of it I am Value Decomposition is like a meat grinder. The data you put in it, this is a different kind of meat. If we take the covariance matrix, put it in our meat grinder decomposition, we get out principal components. If we take a contingency table, we have correspondence analysis. If we take a distance matrix, we get multidimensional scaling, discriminant analysis, just Eigen value Eigen vector decomposition,

Doug Steinley 47:23
when you have a categorical outcome, if I were to be able to have my way, I would have matrix algebra via one of the big components that everybody needs to have a strong understanding. That's all we work with. And then everything else falls from that.

Patrick 47:35
When you think about this rather esoteric concept of eigenvalues and eigenvectors is we use this all the time, whether you're aware of it or not. And not only is it a workhorse for us, but for me, Greg, I think it's a way of thinking about our data, your bouquets of actors, right? By the way, did your prom date ever show up? I know you were waiting with your birthday, I

Greg 48:02
go to my prom. All right, so shut up. Okay, I had a scheduling conflict.

Patrick 48:08
So we're all applied researchers, you go out you gather a sample, you obtain measures, you calculate a correlation matrix. And when you look at that correlation matrix, Eigen values and Eigen vectors give you a way of thinking about not just how are these measures related to variables at a time, but how are they structured multivariate Lee, using this underlying concept of these vectors and distances and components, it's a really remarkable way of thinking about our data in a multivariate space. And overlaid

Greg 48:46
on top of all of this matrix algebra, there's a whole calculus of matrix algebra. So for those of you who just aren't completely fulfilled, taking derivatives with respect to single variables, you can take derivatives with respect to vectors with respect to matrices. And as torturous as that might sound. The whole purpose is that these are expandable, these are able to accommodate systems that include any numbers of variables, any numbers of cases. And this is the machinery that's going on under the hood of the things that we do, I think it's nice to be able to know that that machinery is there, that it's accessible, and that it's even visualizable in terms of the ways that we've talked about it.

49:25
Hello, new. I am the architect. I created the matrix harmony of mathematical precision. I've been waiting for you,

Greg 49:36
Patrick. I don't want to do matrix algebra. I hate matrix algebra. Is there something that can take me through all of this matrix manipulation stuff, but that isn't matrix manipulation stuff.

Patrick 49:48
You know, what's funny, folks is you would think that was a gag Greg's doing he's actually asking me that question.

Greg 49:56
I didn't go to prom

Patrick 49:58
boy is all about Bravo's still bitter we could turn to the systematic study of the guinea pig. What we're gonna do next time is the MOT tricks part tray. I don't know.

Greg 50:14
I think we're already in enough legal trouble.

Patrick 50:16
Yeah, that's true.

Greg 50:17
Let's go down a different path. The Right Stuff.

Patrick 50:21
Okay, I was actually gonna cut Greg off and say I'm gonna cut that. But actually, that's really funny. It's so surprising. We're at seven minutes. It's the first funny thing you've said. Next, you're gonna do the right stuff. But with the double you. We're going to pick up sumo right and path tracing and how with a couple of tracing rules, we can do the covariance algebra, but with a path diagram, it is mind numbingly cool. It's beautiful. That will be next time on quantitated.

MovieClip 50:56
You must be able to see it, Mr. Anderson. You must know it by now. You can't win. It's pointless to keep fighting. Why Mr. Andersen why why do you persist? Because I choose to.

Greg 51:08
Thanks, everybody. Bye bye. Thanks so much for joining us. Don't forget to tell your friends to subscribe to

51:15
anyone home. Hey, Jeffy. Great to see you. It's great to be back.

Greg 51:20
I assumed everything would be fine. Given your

Jiffy 51:22
very particular set of skills. Oh yeah. Ethan sold me to someone who locked me in a cage. But whatever. Seriously, how'd you get out? Well, gosh, you know, lemurs like me have two tongues. Do I just use them to pick the lock? I'm quite gifted lingually Okay,

Greg 51:38
so where have you been since then? Well,

Jiffy 51:40
I was in New York at two I document meeting Broadway shows. I went to the top of the Empire State Building, Metropolitan Museum. I visited my favorite pack banquet in Central Park. Tons of cool stuff fun. And one of my dear best Craig threw in do you know Craig?

Greg 51:58
Craig Rodriguez SE has at University of Michigan. Yes.

Jiffy 52:01

He came to visit can we went to tons of clubs. Totally. Shut it down. Wow. Sounds amazing. Well, I'm just glad to have you back. Let's catch up a little bit later. I'm just closing out the episode. Oh, was this the second matrix one. I listened to the first one while I was getting a Swedish massage in the village. Yep, it's the second one. Do you mind if I do the closing? I might be a little rusty. But okay. Be my guest. Thank you so much for joining the quanta dudes. Don't forget to tell your friends to subscribe on Apple podcast, Spotify or wherever they go to get stuff too little to wile out for a run only to have people crash over to the other side of the street. When you start laughing out loud. We're talking to you Alexandra Earhart, nice. You can also follow us on Twitter where we are at quietude pod and visit our revamped website quantity pod.org where you can leave us a message find organized playlist and show notes. Listen to past episodes and other fun stuff. Is it true to their transcripts or now? There's some translation glitches that we have to iron out but yeah, we're trying it out cool. And finally, you can get amazing quantity merch like shirts, mugs and no pads even stuff with my picture on it from red bubble comm where All proceeds go to Donors Choose that org to help support low income schools. You've been listening to quietude the podcast that well better than being sold by one of your co interns and then getting locked in a cage Hello episodes about matrices. Today's episode has been sponsored by multi colinearity or as our new transcript translator calls it before editing multi colon rarity and by Eigen decomposition also known as I can decompositions

53:41

and finally by generalized indices of variability or generalized indices the fair hillbilly, this is most definitely not NP our thesis most F in Italy not and P R.