Cross Validated

Making sense of principal component analysis, eigenvectors & eigenvalues

938

1267

In today's pattern recognition class my professor talked about PCA, eigenvectors and eigenvalues.

I understood the mathematics of it. If I'm asked to find eigenvalues etc. I'll do it correctly like a machine. But I didn't **understand** it. I didn't get the purpose of it. I didn't get the feel of it.

I strongly believe in the following quote:

> You do not really understand something unless you can explain it to your grandmother. -- Albert Einstein

Well, I can't explain these concepts to a layman or grandma.

1. Why PCA, eigenvectors & eigenvalues? What was the *need* for these concepts?
2. How would you explain these to a layman?

pca intuition eigenvalues

edited Feb 27 at 17:11
OTStats 145 1 9

asked Sep 15 '10 at 20:05
claws 4,874 3 12 10

94 Good question. I agree with the quote as well. I believe there are many people in statistics and mathematics who are highly intelligent, and can get very deep into their work, but don't deeply understand what they are working on. Or they do, but are incapable of explaining it to others.I go out of my way to provide answers here in plain English, and ask questions demanding plan English answers. – Neil McGuigan Sep 15 '10 at 21:43

7 This was asked on the Mathematics site in July, but not as well and it didn't get many answers (not surprising, given the different focus there). math.stackexchange.com/questions/1146/… – whuber ♦ Sep 16 '10 at 5:03

6 Similar to explanation by Zuur et al in Analyzing ecological data where they talk about projecting your hand on an overhead projector. You keep rotating your hand so that the projection on the wall looks pretty similar to what you think a hand should look like. – Roman Luštrik Sep 16 '10 at 9:00

12 This question lead me to a good paper, and even though I think that is a great quote it is not from Einstein. This is a common misattribution, and the more likely original quote is probably this one from Ernest Rutherford who said, "If you can't explain your physics to a barmaid it is probably not very good physics." All the same thanks for starting this thread. – gavaletz Apr 15 '13 at 15:03

22 Alice Calaprice, *The ultimate quotable Einstein*, Princeton U.P. 2011 flags the quotation here as one of many "Probably not by Einstein". See p.482. – Nick Cox Jun 20 '13 at 10:01

28 Answers

1229

+50

Imagine a big family dinner, where everybody starts asking you about PCA. First you explain it to your great-grandmother; then to you grandmother; then to your mother; then to your spouse; finally, to your daughter (who is a mathematician). Each time the next person is less of a layman. Here is how the conversation might go.

Great-grandmother: I heard you are studying "Pee-See-Ay". I wonder what that is...

You: Ah, it's just a method of summarizing some data. Look, we have some wine bottles standing here on the table. We can describe each wine by its colour, by how strong it is, by how old it is, and so on (see this very nice visualization of wine properties taken from here). We can compose a whole list of different characteristics of each wine in our cellar. But many of them will measure related properties and so will be redundant. If so, we should be able to summarize each wine with fewer characteristics! This is what PCA does.

Grandmother: This is interesting! So this PCA thing checks what characteristics are redundant and discards them?

You: Excellent question, granny! No, PCA is not selecting some characteristics and discarding the others. Instead, it constructs some *new* characteristics that turn out to summarize our list of wines well. Of course these new characteristics are constructed using the old ones; for example, a new characteristic might be computed as wine age minus wine acidity level or some other combination like that (we call them *linear combinations*).

In fact, PCA finds the best possible characteristics, the ones that summarize the list of wines as well as only possible (among all conceivable linear combinations). This is why it is so useful.

Mother: Hmmm, this certainly sounds good, but I am not sure I understand. What do you actually mean when you say that these new PCA characteristics "summarize" the list of wines?

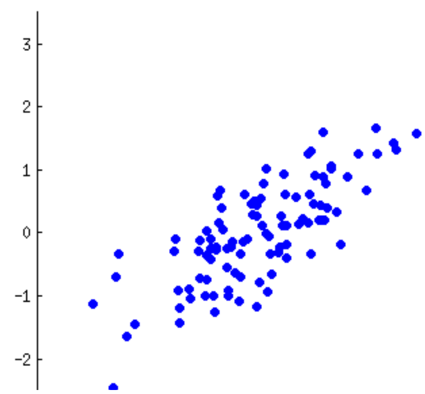
You: I guess I can give two different answers to this question. First answer is that you are looking for some wine properties (characteristics) that strongly differ across wines. Indeed, imagine that you come up with a property that is the same for most of the wines. This would not be very useful, wouldn't it? Wines are very different, but your new property makes them all look the same! This would certainly be a bad summary. Instead, PCA looks for properties that show as much variation across wines as possible.

The second answer is that you look for the properties that would allow you to predict, or "reconstruct", the original wine characteristics. Again, imagine that you come up with a property that has no relation to the original characteristics; if you use only this new property, there is no way you could reconstruct the original ones! This, again, would be a bad summary. So PCA looks for properties that allow to reconstruct the original characteristics as well as possible.

Surprisingly, it turns out that these two aims are equivalent and so PCA can kill two birds with one stone.

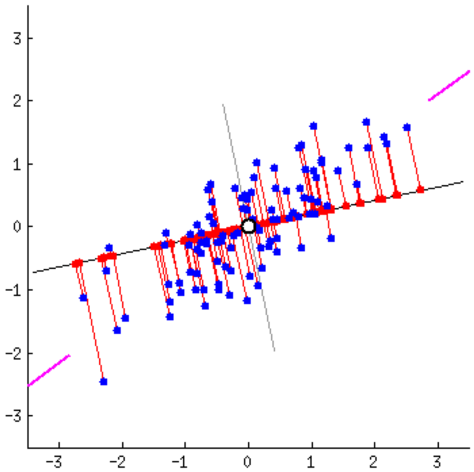
Spouse: But darling, these two "goals" of PCA sound so different! Why would they be equivalent?

You: Hmmm. Perhaps I should make a little drawing *(takes a napkin and starts scribbling)*. Let us pick two wine characteristics, perhaps wine darkness and alcohol content -- I don't know if they are correlated, but let's imagine that they are. Here is what a scatter plot of different wines could look like:



Each dot in this "wine cloud" shows one particular wine. You see that the two properties (x and y on this figure) are correlated. A new property can be constructed by drawing a line through the center of this wine cloud and projecting all points onto this line. This new property will be given by a linear combination $w_1 x + w_2 y$, where each line corresponds to some particular values of w_1 and w_2.

Now look here very carefully -- here is how these projections look like for different lines (red dots are projections of the blue dots):



As I said before, PCA will find the "best" line according to two different criteria of what is the "best". First, the variation of values along this line should be maximal. Pay attention to how the "spread" (we call it "variance") of the red dots changes while the line rotates; can you see when it reaches maximum? Second, if we reconstruct the original two characteristics (position of a blue dot) from the new one (position of a red dot), the reconstruction error will be given by the length of the connecting red line. Observe how the length of these red lines changes while the line rotates; can you see when the total length reaches minimum?

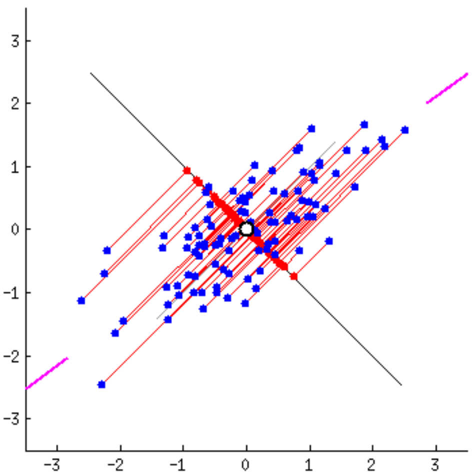
If you stare at this animation for some time, you will notice that "the maximum variance" and "the minimum error" are reached at the same time, namely when the line points to the magenta ticks I marked on both sides of the wine cloud. This line corresponds to the new wine property that will be constructed by PCA.

By the way, PCA stands for "principal component analysis" and this new property is called "first principal component". And instead of saying "property" or "characteristic" we usually say "feature" or "variable".

Daughter: Very nice, papa! I think I can see why the two goals yield the same result: it is essentially because of the Pythagoras theorem, isn't it? Anyway, I heard that PCA is somehow related to eigenvectors and eigenvalues; where are they on this picture?

You: Brilliant observation. Mathematically, the spread of the red dots is measured as the average squared distance from the center of the wine cloud to each red dot; as you know, it is called the *variance*. On the other hand, the total reconstruction error is measured as the average squared length of the corresponding red lines. But as the angle between red lines and the black line is always 90°, the sum of these two quantities is equal to the average squared distance between the center of the wine cloud and each blue dot; this is precisely Pythagoras theorem. Of course this average distance does not depend on the orientation of the black line, so the higher the variance the lower the error (because their sum is constant). This hand-wavy argument can be made precise (see here).

By the way, you can imagine that the black line is a solid rod and each red line is a spring. The energy of the spring is proportional to its squared length (this is known in physics as the Hooke's law), so the rod will orient itself such as to minimize the sum of these squared distances. I made a simulation of how it will look like, in the presence of some viscous friction:



Regarding eigenvectors and eigenvalues. You know what a *covariance matrix* is; in my example it is a 2×2 matrix that is given by

$$\begin{pmatrix} 1.07 & 0.63 \\ 0.63 & 0.64 \end{pmatrix}.$$

What this means is that the variance of the x variable is 1.07, the variance of the y variable is 0.64, and the covariance between them is 0.63. As it is a square symmetric matrix, it can be diagonalized by choosing a new orthogonal coordinate system, given by its eigenvectors (incidentally, this is called *spectral theorem*); corresponding eigenvalues will then be located on the diagonal. In this new coordinate system, the covariance matrix is diagonal and looks like that:

$$\begin{pmatrix} 1.52 & 0 \\ 0 & 0.19 \end{pmatrix},$$

meaning that the correlation between points is now zero. It becomes clear that the variance of any projection will be given by a weighted average of the eigenvalues (I am only sketching the intuition here). Consequently, the maximum possible variance (1.52) will be achieved if we simply take the projection on the first coordinate axis. It follows that the direction of the first principal component is given by the first eigenvector of the covariance matrix. (More details here.)

You can see this on the rotating figure as well: there is a gray line there orthogonal to the black one; together they form a rotating coordinate frame. Try to notice when the blue dots become uncorrelated in this rotating frame. The answer, again, is that it happens precisely when the black line points at the magenta ticks. Now I can tell you how I found them: they mark the direction of the first eigenvector of the covariance matrix, which in this case is equal to $(0.81, 0.58)$.

Per popular request, I shared the Matlab code to produce the above animations.

edited Mar 14 '18 at 13:03
djvg
103 4

answered Mar 6 '15 at 0:30
amoeba
64.3k 16 214 273

75 +1 Nice tale and illustrations. `...then to your mother; then to your wife; finally, to your daughter (who is a mathematician)...` I'd continue: and after the dinner - to yourself. And here you suddenly got stuck... – ttnphns Mar 8 '15 at 21:47

65 I absolutely love the illustrations you make for these answers. – shadowtalker Mar 10 '15 at 16:00

58 +1 The animation is absolutely amazing. This should be the top-answer. – SmallChess Jun 6 '15 at 11:40

52 I normally just browse through Cross Validated to read up on things, but I've never had reason to create an account... mainly because the kinds of questions here are outside of my expertise and I can't really answer any. I usually am on StackOverflow only and I've been on the StackExchange network for about a year now. However, I've only decided to create an account today primarily to upvote your post. This is probably the best exposition of PCA that I've ever read, and I've read many. Thank you for this wonderful post - the excellent storytelling, the graphics, and it's so easy to read! +1 – rayryeng Aug 23 '15 at 21:54

13 Note for myself: my answer currently has 100 upvotes, JDLong's one has 220 upvotes; if we assume constant growth then mine has 100 upvotes/year and his has 40 upvotes/year. Or rather 55/year if computed since it passed 100 upvotes [got a golden badge] in Jan 2014. This means that I will catch up in 2.5--3 years, around the end of 2018. Let's see :-) – amoeba Mar 4 '16 at 18:29

369

+150

The manuscript "A tutorial on Principal Components Analysis" by Lindsay I Smith really helped me grok PCA. I think it's still too complex for explaining to your grandmother, but it's not bad. You should skip first few bits on calculating eigens, etc. Jump down to the example in chapter 3 and look at the graphs.

I have some examples where I worked through some toy examples so I could understand PCA vs. OLS linear regression. I'll try to dig those up and post them as well.

edit: You didn't really ask about the difference between Ordinary Least Squares (OLS) and PCA but since I dug up my notes I did a blog post about it. The very short version is OLS of y ~ x minimizes error perpendicular to the independent axis like this (yellow lines are examples of two errors):