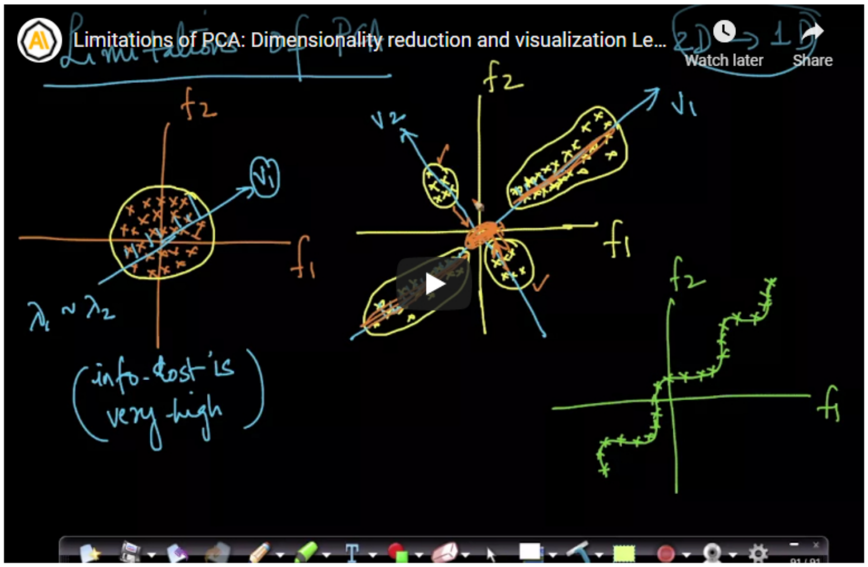
Limitations of PCA

COMPLETED

Instructor: Applied AI Course **Duration:** 5 mins



Prev
Visualize MNIST dataset

Next
PCA Code example

226 Comment(s)

search comments Search

Leave a response

Format

Submit

sanjeev kumar 101 Votes

Advantages and Disadvantages of Principal Component Analysis in Machine Learning

Principal Component Analysis (PCA) is a statistical techniques used to reduce the dimensionality of the data (reduce the number of features in the dataset) by selecting the most important features that capture maximum information about the dataset.

The features are selected on the basis of variance that they cause in the output. Original features of the dataset are converted to the Principal Components which are the linear combinations of the existing features. The feature that causes highest variance is the first Principal Component. The feature that is responsible for second highest variance is considered the second Principal Component, and so on.

In simple words, Principal Component Analysis is a method of extracting important features (in the form of components) from a large set of features available in a dataset.

PCA finds the directions of maximum variance in high-dimensional data and project it onto a smaller

- Plotting for exploratory data analysis (EDA)
- Linear Algebra
- Probability and Statistics
- Interview Questions on Probability and statistics
- Dimensionality reduction and Visualization:
- PCA(principal component analysis)
  - 6.1 Why learn PCA? — 4 min
  - 6.2 Geometric intuition of PCA — 14 min
  - 6.3 Mathematical objective function of PCA — 13 min
  - 6.4 Alternative formulation of PCA: Distance minimization — 10 min
  - 6.5 Eigen values and Eigen vectors (PCA): Dimensionality reduction — 23 min
  - 6.6 PCA for Dimensionality Reduction and Visualization — 10 min
  - 6.7 Visualize MNIST dataset — 5 min
  - 6.8 Limitations of PCA — 5 min
  - 6.9 PCA Code example — 19 min
  - 6.10 PCA for dimensionality reduction (not-visualization) — 15 min
- (t-SNE)T-distributed Stochastic Neighbourhood Embedding
- Interview Questions on Dimensionality Reduction
- Module 2: Live Sessions

dimensional subspace while retaining most of the information. By projecting our data into a smaller space, we're reducing the dimensionality of our feature space.

Following are some of the advantages and disadvantages of Principal Component Analysis:

Advantages of Principal Component Analysis

1. Removes Correlated Features: In a real world scenario, this is very common that you get thousands of features in your dataset. You cannot run your algorithm on all the features as it will reduce the performance of your algorithm and it will not be easy to visualize that many features in any kind of graph. So, you MUST reduce the number of features in your dataset.

You need to find out the correlation among the features (correlated variables). Finding correlation manually in thousands of features is nearly impossible, frustrating and time-consuming. PCA does this for you efficiently.

After implementing the PCA on your dataset, all the Principal Components are independent of one another. There is no correlation among them.

2. Improves Algorithm Performance: With so many features, the performance of your algorithm will drastically degrade. PCA is a very common way to speed up your Machine Learning algorithm by getting rid of correlated variables which don't contribute in any decision making. The training time of the algorithms reduces significantly with less number of features.

So, if the input dimensions are too high, then using PCA to speed up the algorithm is a reasonable choice.

3. Reduces Overfitting: Overfitting mainly occurs when there are too many variables in the dataset. So, PCA helps in overcoming the overfitting issue by reducing the number of features.

4. Improves Visualization: It is very hard to visualize and understand the data in high dimensions. PCA transforms a high dimensional data to low dimensional data (2 dimension) so that it can be visualized easily.

We can use 2D Scree Plot to see which Principal Components result in high variance and have more impact as compared to other Principal Components.

Even the simplest IRIS dataset is 4 dimensional which is hard to visualize. We can use PCA to reduce it to 2 dimension for better visualization.

Consider a situation where we have 50 features (p = 50). There can be p(p-1)/2 scatter plots i.e. 1225 plots possible to analyze the variable relationships. It would be a tedious job to perform exploratory analysis on this data. That is why, we have to use PCA to get rid of this problem.

Disadvantages of Principal Component Analysis

1. Independent variables become less interpretable: After implementing PCA on the dataset, your original features will turn into Principal Components. Principal Components are the linear combination of your original features. Principal Components are not as readable and interpretable as original features.

2. Data standardization is must before PCA: You must standardize your data before implementing PCA, otherwise PCA will not be able to find the optimal Principal Components.

For instance, if a feature set has data expressed in units of Kilograms, Light years, or Millions, the variance scale is huge in the training set. If PCA is applied on such a feature set, the resultant loadings for features with high variance will also be large. Hence, principal components will be biased towards features with high variance, leading to false results.

Also, for standardization, all the categorical features are required to be converted into numerical features before PCA can be applied.

PCA is affected by scale, so you need to scale the features in your data before applying PCA. Use StandardScaler from Scikit Learn to standardize the dataset features onto unit scale (mean = 0 and standard deviation = 1) which is a requirement for the optimal performance of many Machine Learning algorithms.

3. Information Loss: Although Principal Components try to cover maximum variance among the features in a dataset, if we don't select the number of Principal Components with care, it may miss some information as compared to the original list of features.

Aug 17, 2019 17:27 PM

dhilip vasanth
thanks sanjeev kumar well suggest on the pca i got clear intuition behind it.

Reply

Mar 13, 2020 19:38 PM

Omkar Chavan
i think overfittin is not related to too many variables its just related to not findin generalized pattern in those variables, infact pca is reducin features so it might overfit.

Reply

Aug 26, 2020 10:42 AM

appliedai course
yes, correct. reducing number of features does not mean we are reducing overfitting

Reply

Aug 26, 2020 12:06 PM

Prashant Manoj Mishra
Hey I don't agree with the fact that PCA definitely helps to reduce overfitting. Here we are talking about dimensionality reduction and not feature selection. Feature selection does reduce overfitting. PCA might work but it isn't a good way to address the problem of overfitting.
https://www.coursera.org/learn/machine-learning/lecture/RBqQl/advice-for-applying-pca
You can refer 6:55 of the above video.
Please correct me if my understanding is wrong.

Reply

Mar 22, 2020 20:48 PM

Ritwik
You're right. In Andrew NG's words, PCA does not necessarily reduce overfitting and shouldn't be used for dealing with overfitting. To deal with overfitting, it is much better to use regularization. Even if in some cases, PCA helps you avoid over fitting, it is not a good practice! I was so hoping to see a comment like yours!

Reply

Apr 14, 2020 23:19 PM

Shyama Maria Kurian
Hi Team,

The number of features is equal to the dimension of the data if I am not wrong. So if the number of features is 5, then it is a 5-dimensional dataset.

> So won't feature selection and dimensionality reduction the same? Or is it like, PCA is a dimensional reduction method, as we find new features which best describe the output, while the new feature contains the maximum information of all the features?
>
> Regarding the summary by Sanjeev Kumar, "Removes Correlated Features", by this, does it mean the new feature created helps in reducing correlated features. Say we plot a graph in a 2D space and we want to convert it into a 1D space. We plot heights(x-axis) (feature 1) and weights(y-axis) (feature 2) and we get a positive correlation. So the new feature, *feature 1'* is used instead of heights and weights which are correlated

Thank You

Reply

Dec 18, 2020 14:52 PM

team aaic
number of features = dimensionality of data
1. for feature selection vs dimesionality reduction: refer this
2. yes, correct

Reply

Dec 18, 2020 18:50 PM

Shyama Maria Kurian
Hi Team,

Both Feature Selection and Dimensionality reduction help in reducing the number of features. Say there are 5 features and we want only 3 features for analysis (5D -> 3D)

Feature Selection: Drops 2 features based on certain conditions like multicollinearity, low variation.

Dimensionality Reduction: Creates 5 new features, and then selects the top 3 features which have high variance.

Can you check if my understanding is right?

Thank You

Reply

Dec 19, 2020 08:47 AM

team aaic

Yes, correct.

Reply

Dec 19, 2020 11:34 AM

Shyama Maria Kurian

Thank you, Team

Reply

Dec 20, 2020 10:57 AM

Uttam Dey

Thank you Sanjeev Kumar for the this short and nice conclusion on PCA and its advantages and disadvantages.

Reply

Apr 01, 2020 16:46 PM

Kuldeep Pal

http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of_4.html

Reply

Apr 15, 2020 23:18 PM

CR

source of above post : http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of_4.html

Reply

Jul 02, 2020 22:20 PM

NAVEEN SHRIVASTAVA

Agreed with the explanation except the overfitting which better deal with regularisation instead of PCA

Reply

Jul 29, 2020 20:44 PM

Ranjith

16 Votes

A simple thought is that if for example there are 2 3-D vectors with co-ordinates [1,2,3] and [1,2,4] and we remove the third dimension (like, the coordinates 3 and 4 for the third dimension), and project them to the remaining 2 Dimensions they overlap and we can't separate them. The problem is when we have many such points when we reduce dimensions. Correct?

Reply

Jun 27, 2019 09:23 AM

AppliedAI Course

Yes, you are correct.

Reply

Jun 27, 2019 09:36 AM

shashank Desai

Hi,

Here if we are converting 2D to 3D we are getting two prencipal components which have maximal

Here if we are converting 2D to 3D we are getting two principal components which have maximal variance.Those principal componenets are liner compinations of the features/dimesions.Hence the above logic is not valid.

Please corect me if i am wrong.

Reply

Sep 23, 2020 09:17 AM

team aaic

No, it is not linear combination:

To transform x_q from 3d to 2d

$$x_q^T = \begin{pmatrix} x_{q1} & x_{q2} & x_{q3} \end{pmatrix}_{(1,3)} \begin{pmatrix} \Uparrow & \Uparrow \\ v_1 & v_2 \\ \Downarrow & \Downarrow \end{pmatrix}_{(3,2)}$$

Reply

Sep 23, 2020 09:28 AM

---

vashist narayan singh

5 Votes

did not get the last sine way structure whats wrong if we loose that sin wave structure if we reduce the dimension to 1D from 2D

Reply

Apr 30, 2019 16:40 PM

Applied AI Course Team1

If we will reduce the dimension our sine wave property will get diminish and will not able to guess about the form of data we have.Sinusoidal nature will be deleted from the data hence this is huge loss of information from our dataset.

Reply

Apr 30, 2019 23:41 PM

Rachna Dhanraj

Hello Team,

I too have a question on same lines.

I thought that only the data-points matter, and once we have all the data-points collected(that is if we retain 100% variance), we do not lose anything.

But here you are saying that sinusoidal nature is lost, I am a bit confused.

Like is there is hierarchy here too, that:-

1. We need to collect all the data points projected
2. We should try not to lose the fundamental property(here it is its sine nature)
3. and so on.

Thanks in advance.

Reply

May 09, 2019 12:47 PM

Applied AI Course Team1

See when we applied PCA on the sinusoidal data we have the maximum variance in one dimension here(lets suppose) the projected points in the direction of maximum variance is the best solution to the PCA.Now we have the single dimensional data and we cant say that above data has sinusoidal nature in past.SO sinusoidal property sometimes helpfull in deciding the type of the model to be used etc.I hope you now get the idea of how sinusoidal nature is lost and why.

Reply

May 10, 2019 00:28 AM

Narayanam Kundhan

I think in this case the classification is still possible but in case of prediction model the accuracy suffers

Reply

Jan 02, 2020 23:41 PM

Applied AI Course Team1

Yes we can have that as a solution no issues but error can be high surely

Reply

Jan 03, 2020 22:13 PM

Chakravarthy Karri

3 Votes

Can you please explain why the third fig shown in this video considered as limitation? Though we loose the shape of the data, we still have our projections clearly separated and I hope this is what we are trying to achieve with PCA.

Reply | Jul 24, 2019 23:21 PM

AppliedAICourse
Well, the structure of data as well is important. We are loosing the information of amplitude of those waves here. So it is considered as limitation.

Reply | Jul 25, 2019 08:04 AM

Sumit Kumar
Why is the structure important?Can you please give one real life example to elaborate the same?

Reply | Dec 04, 2019 17:27 PM

AppliedAI Course
Please refer to this comment.

Reply | Dec 05, 2019 07:27 AM

---

raghavendra kanike — 3 Votes
so after this explanation, i got to understand only thing. That is, we can apply PC for only linearly saparable data.
is that correct?
if not, for what else shpes we can apply PCA?

Reply | May 20, 2019 22:20 PM

AppliedAI
Here it means that if you have some of the variables in your dataset that are linearly correlated, PCA can find directions that represents your data, but if the data is not linearly correlated (f.e. in spiral, where x=t*cos(t) and y =t*sin(t)), PCA is not enough.

Reply | May 21, 2019 15:13 PM

lokesh madasu
What is linearly correlated?Can u please explain

Reply | Jun 28, 2019 20:58 PM

Applied AI Course Team1
Leanr correlation means if any feature A can be written as bx + c where b and c are other features.

Reply | Jun 29, 2019 00:46 AM

Faisal Rasheed — 2 Votes
Cant we have all these informations in some written or textual form??

Reply | Apr 01, 2019 14:11 PM

Applied AI Course Team1
One of the best remedies will be to take notes of these sessions as they will help you in future references.

Reply | Apr 01, 2019 16:58 PM

Faisal Rasheed
Those notes have only figures not explanation in words

Reply | Apr 01, 2019 17:07 PM

AppliedAICourse
Sorry, we don't have all these in textual form, we suggest you prepare the running notes by yourself.

Reply | Apr 02, 2019 05:18 AM

Abhinav kumar — 1 Votes
Hi Sir,

As we know to Perform PCA we need to normalize the data.

So we should perform Standardization (mean 0, SD 1) or Normaization (range 0,1) in PCA?

Reply

Nov 13, 2020 16:44 PM

team aaic

We need to standardize the data to mean 0 and standard deviation 1.

Reply

Nov 13, 2020 17:02 PM

Abhinav kumar

Hi Sir,
What if we perform Normalization (range 0,1) ?

Reply

Nov 15, 2020 09:42 AM

team aaic

While formulating the PCA, we used covariance. Here, we assumed zero mean. So, if we normalization we can't ensure zero mean.

Reply

Nov 15, 2020 10:11 AM

Jayakrishnan

1 Votes

hi team,
after pca will the correlation among feature be removed?

Reply

Mar 26, 2020 18:44 PM

AppliedAI Course

Yes, correct. Refer to first point here.

Reply

Mar 26, 2020 19:00 PM

Vinay Sawant

1 Votes

PCA works fairly well for data with linear relaltionships,for non linear ,there is lot of information loss!

Reply

Feb 25, 2020 17:19 PM

AppliedAI Course

Yes, correct.

Reply

Feb 25, 2020 17:24 PM

ANUJ SRIVASTAVA

1 Votes

Are PCA still in use in general mathematics?

Reply

Jul 06, 2019 10:59 AM

AppliedAICourse Team

Yes PCA is one of the most powerful techniques to find the axes with maximal variance.

Reply

Jul 06, 2019 18:25 PM

Terms & Conditions