Revision Questions

COMPLETE

Instructor: Applied AI Course **Duration:** 30 mins

Revision:

What is dimensionality reduction? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2878/what-is-dimensionality-reduction/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

Explain Principal Component Analysis?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2889/geometric-intuition-of-pca/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

Importance of PCA?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2888/why-learn-pca/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

Limitations of PCA?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2894/limitations-of-pca/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

What is t-SNE?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2898/what-is-t-sne/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

What is Crowding problem?https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2901/crowding-problem/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

How to apply t-SNE and interpret its output? https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/2902/how-to-apply-t-sne-and-interpret-its-output/2/module-2-data-science-exploratory-data-analysis-and-data-visualization

Prev

Code example of t-SNE

Next

Questions & Answers

35 Comment(s)

search comments Search

Leave a response

Format

Submit

Ramya Vidiyala *23 Votes*

Here are my answers. Please let me know if anything is wrong or needs to be added.
1 Dimensionality reduction means projecting data to a lower-dimensional space, which makes it easier for the visualization and analysis of data.

2 Explain Principal Component Analysis?
PCA means finding out the components (features) which are effective to the data and discarding the redundant features.

3 Importance of PCA?
With few lines of codes we can reduce the dimensions by a huge number.

4 Limitations of PCA?
PCA does preserve the global direction of the data but not the local, which creates confusion when an overlap of 2 clusters happens after the reduction.

5 What is t-SNE?
t-SNE stands for t-distribution Scholastic Neighbourhood Embedding.
Scholastic – not definite but random probability
Neighborhood – concerned only about retaining the structure of neighborhood points.
Embedding – plotting data into lower dimensions tSNE is the state of the art or one of the best techniques for dimensionality reduction, which is widely used for data visualization.

6 What is Crowding problem?
When a data point, ‘x’ is a neighbor to 2 data points that are not neighboring to each other, this may result in losing the neighborhood of ‘x' with one of the data points as t-SNE is concerned only within the neighborhood zone.

7 How to apply t-SNE and interpret its output?

· There are 3 parameters
a) Steps: number of iterations
b) Perplexity: can be thought of as the number of neighboring points.
c) Epsilon: It is for data visualization and determines the speed which it should be changed.

· Points to remember while performing tSNE are

1. Never stop with a single-step value. Check for various values and take the value at which the plot is stable.

2. With lower perplexity values, we may see a few shapes of clusters. But do not fall into the trap. Try with various Perplexity values ranging from 2 to the number of data points. But, remember a value of 2 or a value equal to a number of data points will lead to no information.

3. Never come to any conclusions with random data.

4. As tSNE is Scholastic, each run may lead to slightly different. However, by setting random_state, this can be solved.

5. tSNE doesn’t preserve the distance between clusters. So, when we have multiple clusters, we might not retain the similar distance between the clusters.

6. tSNE shrinks the widespread data and expands densely packed data. So based on the output, we cannot decide on the cluster size and density/ spread / variance of the clusters.

Reply *May 16, 2020 15:55 PM*

team aaic

Your understanding is correct. You can refer this blog for better understanding of PCA importance.

Reply *May 16, 2020 18:47 PM*

Slim Shady

which blog?

Reply *Jun 27, 2020 21:05 PM*

team aaic

Please refer this blog : https://towardsdatascience.com/pca-clearly-explained-how-when-why-to-use-it-and-feature-importance-a-guide-in-python-7c274582c37e

Reply *Jun 28, 2020 10:53 AM*

Sujit Jena *8 Votes*

What is Dimensionality Reduction?

In machine learning classification problems, there are often too many factors on the basis of which the final classification is done. These factors are basically variables called features. The higher the number of features, the harder it gets to visualize the training set and then work on it. Sometimes, most of these features are correlated, and hence redundant. This is where dimensionality reduction algorithms come into play. Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables. It can be divided into feature selection and feature extraction.

Reply *May 14, 2019 14:04 PM*

AppliedAI Team

what is doubt here?

Reply *May 14, 2019 20:41 PM*

Sujit Jena

I have shared this as an answer , let me know if I am wrong

Reply *May 14, 2019 20:44 PM*

AppliedAI Team

That is right.

Reply *May 14, 2019 21:34 PM*

Chinda Mani Teja Verma *6 Votes*

give some pdf document to the above questions.

Reply *Feb 01, 2019 12:58 PM*

Applied AI Course

Sure, will work on it.

Reply *Feb 02, 2019 08:29 AM*

Debasish Acharya

can we get the pdf?

Reply *Jun 07, 2020 01:02 AM*

Applied_AI

Please drop us a mail at teamappliedaicourse.com

Reply *Jun 07, 2020 12:57 PM*

Sadiva Madaan *4 Votes*

Dimensionality Reduction means projecting a data matrix from a higher dimension to a lower dimension . It basically removes all the redundant features from our data .

PCA helps in finding the most important features . With the help of it we can find eigen values and eigen vectors. They help us to know by how much we have to rotate our axis for maximum variance / information .

Importance of PCA - We can massively reduce the dimensions of our data matrix with a few lines of code.

Limitations of PCA - PCA doesn't work well when we have our data distributed in the form of a circle or in clusters . It preserves the global shape but fails to preserve the local shape .

t-SNE - t - Students t distribution, S - Stochastic (It means that it is not deterministic but is probabilistic), N - Neighborhood (t-SNE main objective is to preserve the structure of neighborhood points), E - Embedding (It means picking up a point from high dimensional space and placing it into lower dimension).

Sometimes it is impossible to preserve the distances in all the neighborhoods . This problem is called Crowding Problem .

How to apply t-SNE -
1) There are two most important parameters - No of steps and Perplexity .
Perplexity means the no of neighborhood points to be preserved . Steps means the no of iterations it should perform .
2) Always run t-SNE with multiple perplexity values .
3) If perplexity = no of data points then it will create a mess.
4) t-SNE never actually replicates the data .

Reply *Aug 12, 2020 22:19 PM*

team aaic

Great summary of both concepts(PCA and t-SNE). Thanks for sharing.

Reply *Aug 12, 2020 22:28 PM*

Ayush Agarwal *3 Votes*

Can you please provide questions like "What change would you do in your model if there is this problem or that,"?

Reply *Mar 02, 2019 12:51 PM*

AppliedAI Team

thanks for your feedback. We will definitely try to create some context based interview question as you suggested and will update these. We have given some questions here.

Reply *Mar 03, 2019 11:54 AM*

Sunney Sood

the link given above does not work..please check and repost. Thanks

Reply *Apr 16, 2019 14:06 PM*

Karthik *1 Votes*

How can we find the best perplexity value ? (Like elbow method of k-means)

Reply *Nov 16, 2019 23:27 PM*

appliedai course

we can fix a large steps like steps=5000 and try out various values of p and take that p as best p when the model stabilises(i.e model does not change as we increase p)

Reply *Nov 17, 2019 05:01 AM*

Subrahmanyam Kesani

As we increase no. of iterations, we will reach a stage where the result will not change further. Is it the same case with perplexity also ? I mean... after certain perplexity...won't the result change anymore by increasing perplexity (using "random_state=0") ? In an example shown in the video "How to apply t-SNE and interpret its output ?", the stabilized cluster points become disarray when perplexity is increased further.
t-SNE undoubtedly is giving beautiful clusters we wish for. But I am doubtful on when to conclude based on perplexity.

Reply *May 11, 2020 16:19 PM*

appliedai course

no if we increase perplexity further cluster points disarray. refer this under 'those hyperparameters really matter" section for constant steps of 5000, perplexity of 30 and 50 are stable but for perplexity 100 results messed up

Reply *May 11, 2020 18:58 PM*

Tanweer Khan *0 Votes*

Hi Team,

Could you please explain this statement "Problems with t-SNE arise when intrinsic dimensions are higher i.e. more than 2-3 dimensions. t-SNE has the tendency to get stuck in local optima"

Thanks.

Reply *Feb 21, 2021 10:02 AM*

team aaic

When intrinsic dimension of data means minimum dimension to represent data is more than 3, then problem of stucking at local optima is problem in t-sne. If you are not familiar with terms like local optima or global optima, you will learn those in optimization chapter.

Reply *Feb 21, 2021 17:06 PM*

Kirandeep Marala *0 Votes*

Hi Team, To this Question Will This Answer Fits:

Hi Team, To this Question Will This Answer Fits.

1. Can we use TSNE for dimensionality reduction i.e convert the data n to d dimension.

Solution:

No, TSNE is Used only to Convert Data From d-dimension to 2-dimension...Is This Correct Answer..Will You PLease Give Correct Answer..Thank YOU..

Reply *Feb 08, 2021 10:39 AM*

team aaic

The main reason that tt-SNE is not used in classification models is that it *does not* learn a function from the original space to the new (lower) dimensional one. As such, when we would try to use our classifier on new / unseen data we will not be able to map / pre-process these new data according to the previous tt-SNE results. Reference

Reply *Feb 08, 2021 10:46 AM*

Amlan Gopal *0 Votes*

Hi Team,

How the performance of PCA and TSNE is measured?

For PCA, is it % of variance explained? What is TSNE? any distance/similaity measure

Reply *Dec 23, 2020 20:30 PM*

team aaic

We can't measure the performance of t-SNE. It is mainly used for visualization purpose.

Reply *Dec 23, 2020 21:01 PM*

Piyush Pandey *0 Votes*

Hi,

In the limitations of PCA, you said that it tries to preserve the global structure, but in the christopher olah's blog he has written that PCA tries to preserve the linearity. Can you comment on that?

Reply *Dec 16, 2020 19:03 PM*

team aaic

PCA is defined as an orthogonal **linear** transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

Reply *Dec 16, 2020 20:08 PM*

Previous 1 2 Next

Our courses

Applied Machine Learning Course

AI/Machine Learning Case Studies

AI Workshop

Contact us

+91 8106-920-029

+91 6301-939-583

(whatsapp business)

More

Success stories

Job Guarantee

Live Sessions

Desktop Application

Student Blogs

Terms & Conditions