



Basics of Machine Learning

Author:
Ramachandran Chandrasekaran
16th July 2021



Introduction – The MIT way!



01





IntroToDeepLearning.com

The Math behind the Magic



02



Introduction to the core mathematical concepts

- *We were introduced to the concepts of Machine Learning and Deep Learning in the previous AI/ML session that we had. Now we delve deep into the core mathematical concepts that form the fundamentals of Machine Learning.*
- *When we speak about the fundamentals of Machine Learning 2 topics come to mind:*
 - *PCA or Principal Component Analysis*
 - *SVD or Singular Value Decomposition.*



A little bit of Linear Algebra



03



Linear Algebra Walk-Through – Linear Spaces

- *In linear algebra, we talk about linear spaces, which are also called vector spaces.*

Hoffman & Kunze's "Linear Algebra"

Sheldon Axler's "Linear Algebra done right"

- *For our case in point, you can consider this linear space or vector space as a multi-dimensional space with each attribute as a separate dimension.*
- *Each linear space has a dimensionality, which is what you might think intuitively as the number of features or the columns/attributes in the data which is minimally required to fully represent the data, but I'll give you a simple example over here to help you in the process of grasping this concept.*



Dimensionality of Vector Spaces

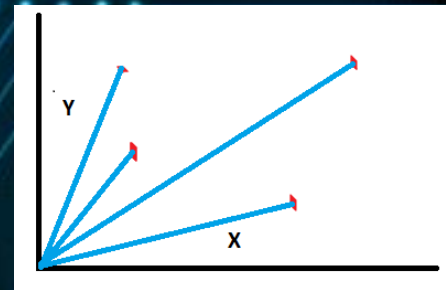
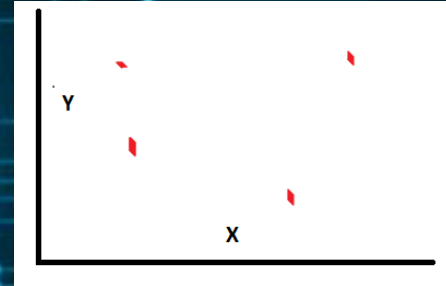
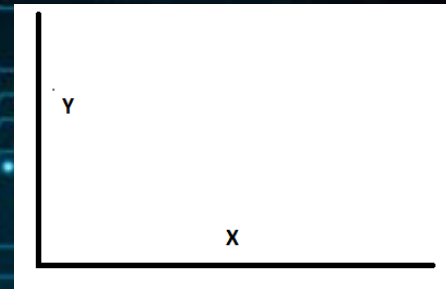


04



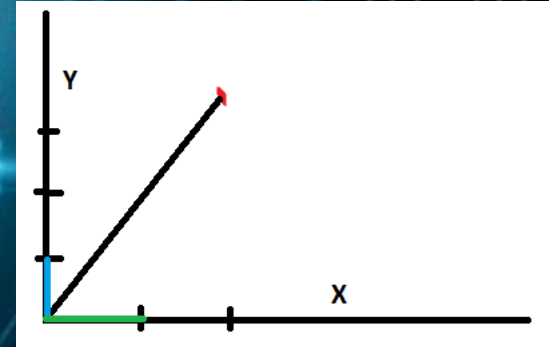
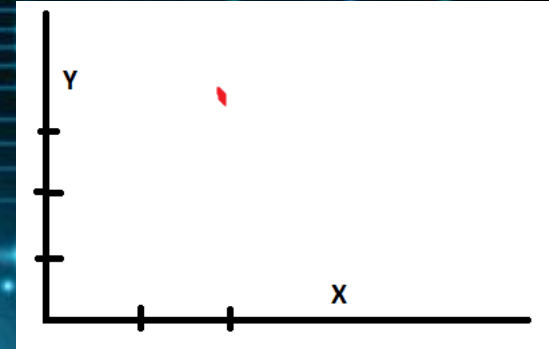
Linear Algebra Walk-Through – Dimensionality

- *Let's consider the two-dimensional linear space of real numbers. This is simply the xy -plane that you're probably used to. Let me ask you a question—can you think of combination of vectors, or points, such that if you take their linear combination, can form the entire 2D space?*
- *A linear combination of two vectors x and y is $ax + by$, for any choice of constants a and b . So basically, here's what I'm asking. Can you find out a set of points, such that if I give you any random point, say $(2, 3.5)$, you'll be able to find some unique constants a and b so that the linear combination of your set of points gives me $(2, 3.5)$?*



Linear Algebra Walk-Through – Dimensionality

- *One easy choice is a unit vector in the x and the y directions. So we could take $(0, 1)$ and $(1, 0)$. So if you gave me $(2, 3.5)$, then $2 * (1, 0) + 3.5 * (0, 1) = (2, 3.5)$, and you see the way we've done the multiplication and addition there.*
- *You should prove to yourself that given any point on the 2D space, you can easily find constants to multiply with this choice of vectors to get that point. We call such a set of points, a **basis**, and the basis is said to **span** the entire 2D linear space.*



Vector Subspaces

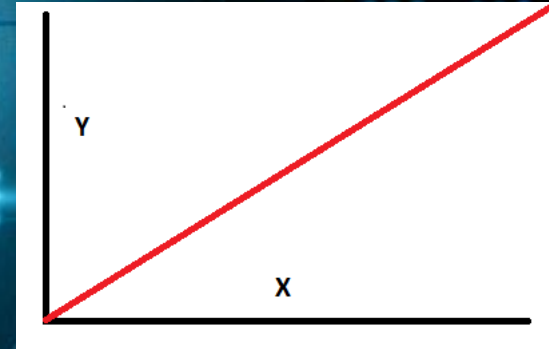
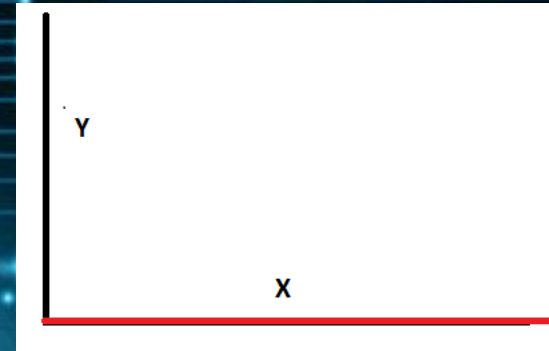


05



Linear Algebra Walk-Through – Subspaces

- *One important concept before we step back into machine learning and away from the mathematical bits is the concept of a subspace. A subspace, also called a manifold, is a subset of a linear space.*
- *For example, we could have a linear subspace for the 2D linear space by restricting the second component to zero.*
- *If you think about it, now you're really just stuck with real numbers since the second component isn't very helpful—and real numbers form a 1D linear space.*
- *Another example would be to put the restriction that both the components must be equal.*



Principal Component Analysis (PCA)

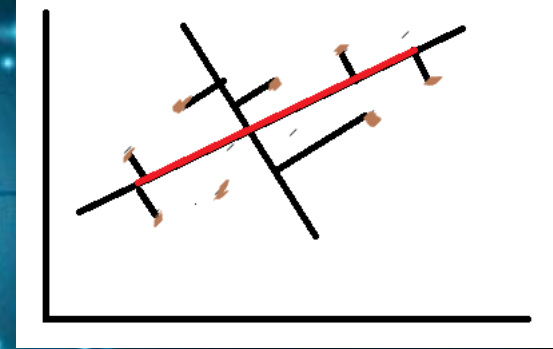


06



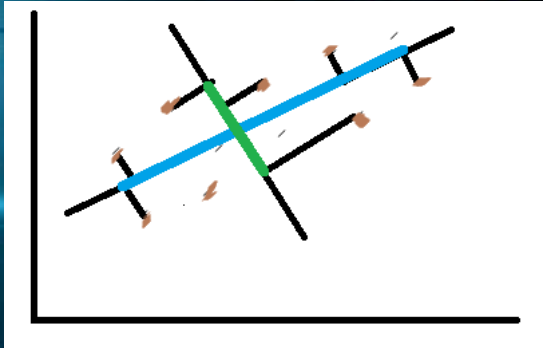
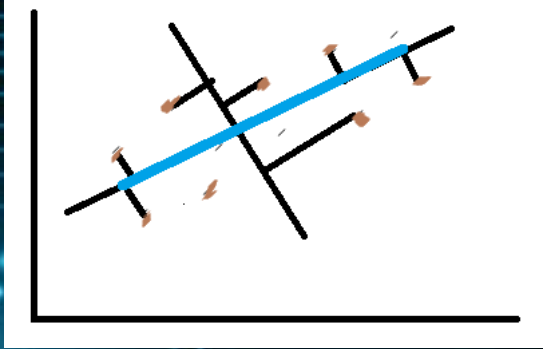
Principal Component Analysis (PCA)

- *PCA is a dimensionality reduction technique which tries to find a linear subspace of the data that you currently have.*
- *So starting from a large set of attributes, you shrink it using PCA to a small set of attributes, which PCA calls as principal components.*
- *PCA assumes that these principal components are orthogonal to each other, and while this may not always be the case, in practice, PCA usually works reasonably well.*



Principal Component Analysis (PCA)

- *PCA transforms a set of vectors in a linear space to an orthogonal basis in a subspace.*
- *The eigenvectors and eigenvalues of a matrix represent the “core” of PCA:*
- *The eigenvectors (principal components) determine the directions of the new feature space, and the eigenvalues determine their magnitude.*



Principal Component Analysis (PCA)

- *An eigenvalue is a number, telling you how much variance there is in the data in that direction.*
- *Simply put, the eigenvalue is a number telling us how spread out the data is on that particular line or eigenvector.*
- *The eigenvector with the highest eigenvalue is therefore the principal component.*



Principal Component Analysis (PCA)

- *An orthogonal basis is simply a basis where each vector is orthogonal to every other.*
- *But that's not the only great thing about PCA. The principal components that it gives you are in descending order of the variance explained.*
- *What does this mean?*

It means that when you take a point in the original linear space and project it to the subspace, then along the first principal component, it will be spread the most as compared to the others.



Uses of Principal Component Analysis



07



Uses of Principal Component Analysis

- *PCA is a dimensionality reduction technique which can be used for various tasks like:*

1. Visualization

For eg: In Stanford, they had data from around 50 points in a monkey's brain and needed a technique to visualize this. So they used PCA to reduce the dimensionality from 50 to a 3D space. This way they were able to visualize what the monkey was thinking in a 3 dimensional graph.

2. Image Compression

We'll later on be working on this example ourselves in today's Hands-on Labs.



Uses of Principal Component Analysis

3. Machine Learning

To make the learning more efficient by performing dimensionality reduction using PCA.

4. Anomaly or Outlier Detection

Not the best anomaly detection algorithm. But, sometimes it's done.

5. Matching/ Distance Calculation:

Compare the projections on the dimensionally-reduced linear subspace for finding similarities and dissimilarities.



Singular Value Decomposition (SVD)



08



Singular Value Decomposition

Singular Value Decomposition (SVD) provides another way to factorize a matrix into singular vectors and singular values.

SVD can be written as:

$$A = U \Sigma V^T$$

where :

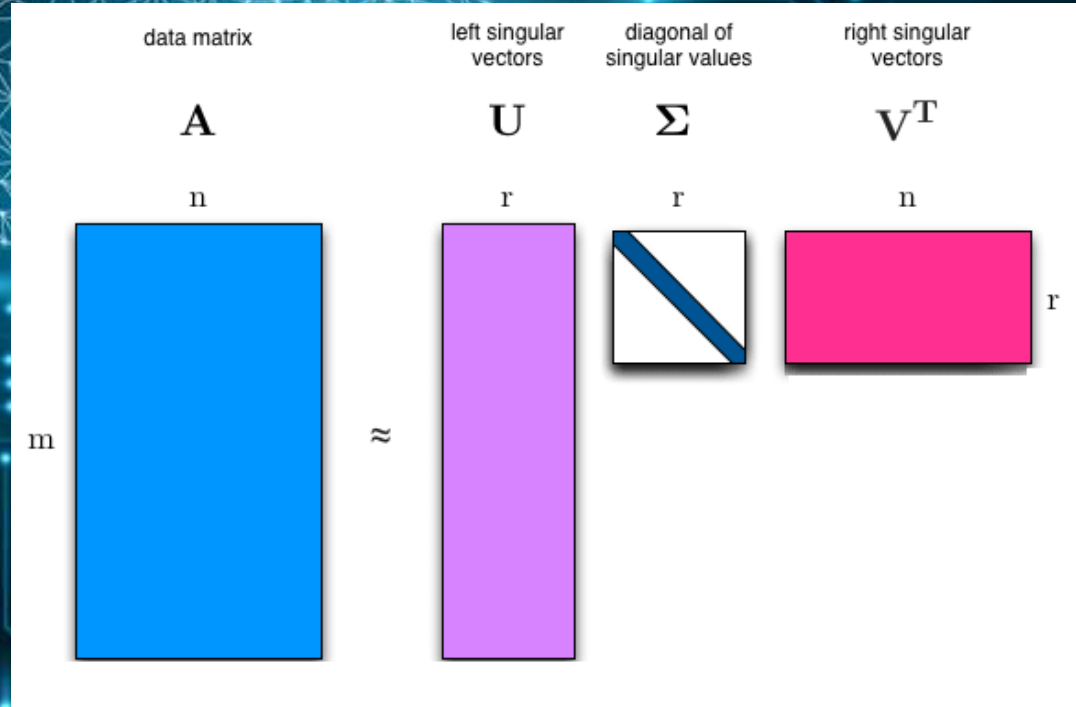
A is an $m \times n$ matrix,

U is an $m \times r$ matrix,

Σ is an $r \times r$ matrix &

V^T is an $r \times n$ matrix.

And r is the Rank of the matrix A.





Let's go to the Hand's on Labs!
HAVE FUN!!!





vodafone
business

Together we can

