

Feature Engineering

Feature Transformation

Feature Construction

Curse of Dimensionality

Missing value imputation

Handling categorical feature

Outlier detection

Feature scaling

Feature Selection

Forward selection

Backward Selection

Feature Extraction

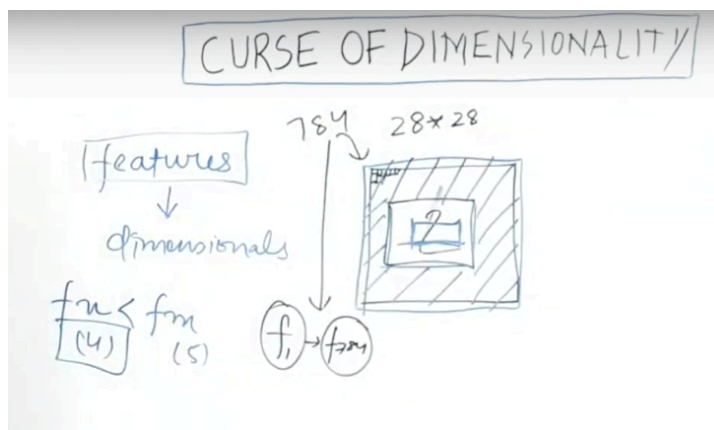
PCA, LDA, tSNE

Curse of Dimensionality is also called Curse of features.

No. of features are imp when we are training a model but it should have some optimal no.

Means if we are training a model with features more than optimal no. then it might be possible that Accuracy of model will be constant or start decreasing.

And if we are training a model with features less than optimal no. then it might be possible that Accuracy of model will be less and not increasing.



In short we can say that with higher dimensions or we can also say that with n. no of features when we train a model it will create 2 problems.

1) Performance of model decreases

2) It will be computationally high.

Note: Variance is proportional to spread it is not exactly spread.

In PCA why Mean Absolute Deviation (MAD) is not using instead of variance? Why we are not taking spread?

Ans:- Because in Mean Absolute Deviation (MAD) we are taking Mod | | and we know that it is not differentiable. And variance is easily differentiable.

Types of Linear Transformation (Stretching, Squeezing, Rotation, Shearing, Reflection, Orthogonal Projection)

Eigen vectors : These are those vector when we apply linear transformation on it, its direction is same ie. It is in the same span but magnitude (scale) is changing. In 2D we will get 2 such vector. In 3D we will get 3 such vector.

After transformation whichever the Eigen vector is available it will acting as axis of transformation. Current orientation of axis (X-Y) is changing on Eigen vector axis of transformation.

Eigen values: It is that value which tell us that by how much quantity Eigen vectors magnitude is changing.

Covariance matrix is that matrix which its diagonal elements tells us that spread of the data on each axis. And along with that its non diagonal elements tells that how the 2 axis are co related with each other.

In short we have to find covariance matrix & then we have to Eigen decomposition of that matrix then we will get Eigen vector and Eigen value then we will select that Eigen vector who's Eigen value is max.

Why we do Eigen decomposition of covariance matrix ?

Ans:- It is giving us largest Eigen vector, largest Eigen vector means who's Eigen value is max so that Eigen vector points into the direction of largest variance of the data.