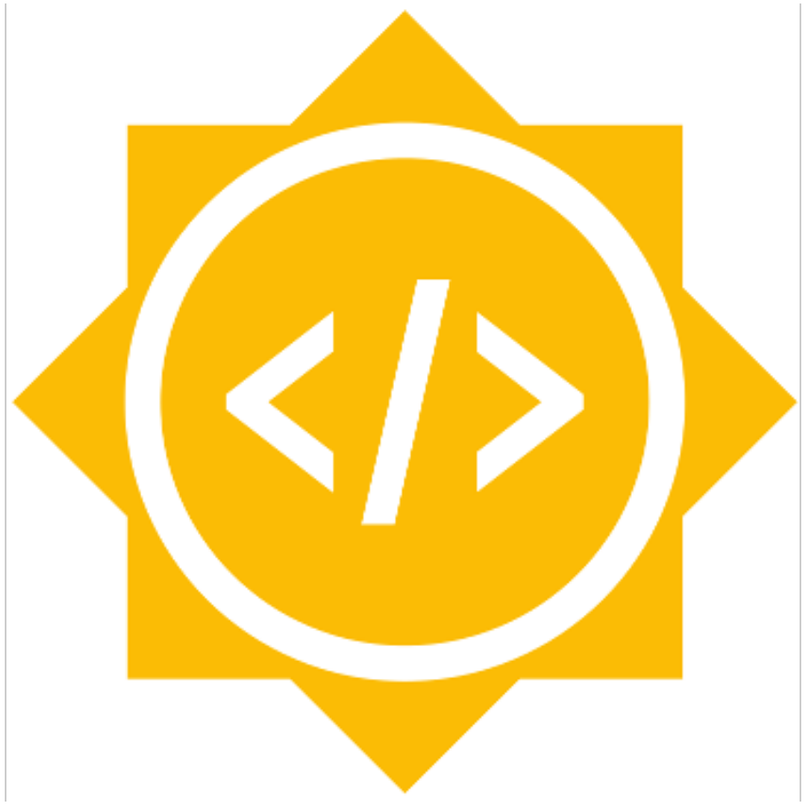


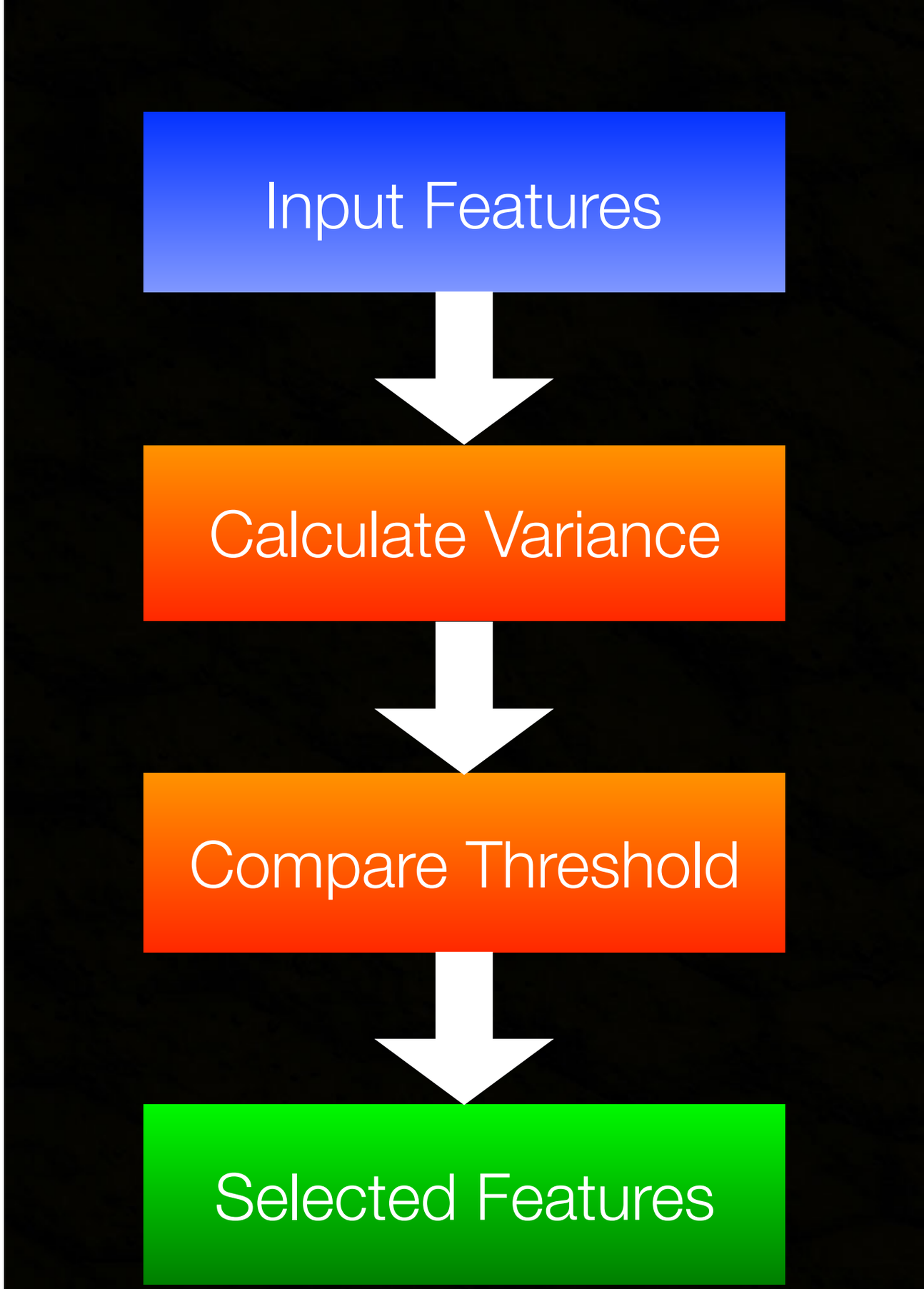
TMVA Project in Machine Learning

Abhinav Moudgil

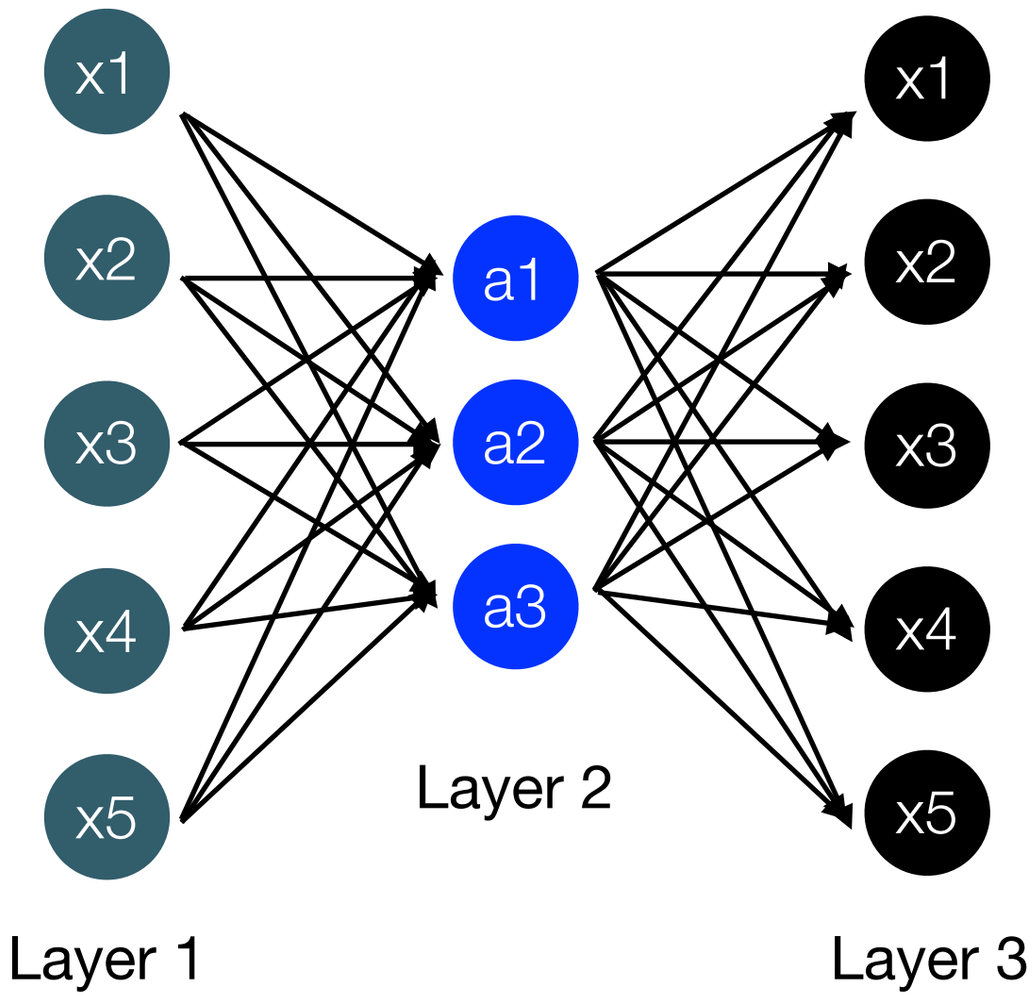
Mentors
Sergei Gleyzer
Omar Zapata
Lorenzo Moneta

Variance Threshold

- Unsupervised feature selection method
- Takes a threshold value as input
- Select features whose variance lie above the threshold value



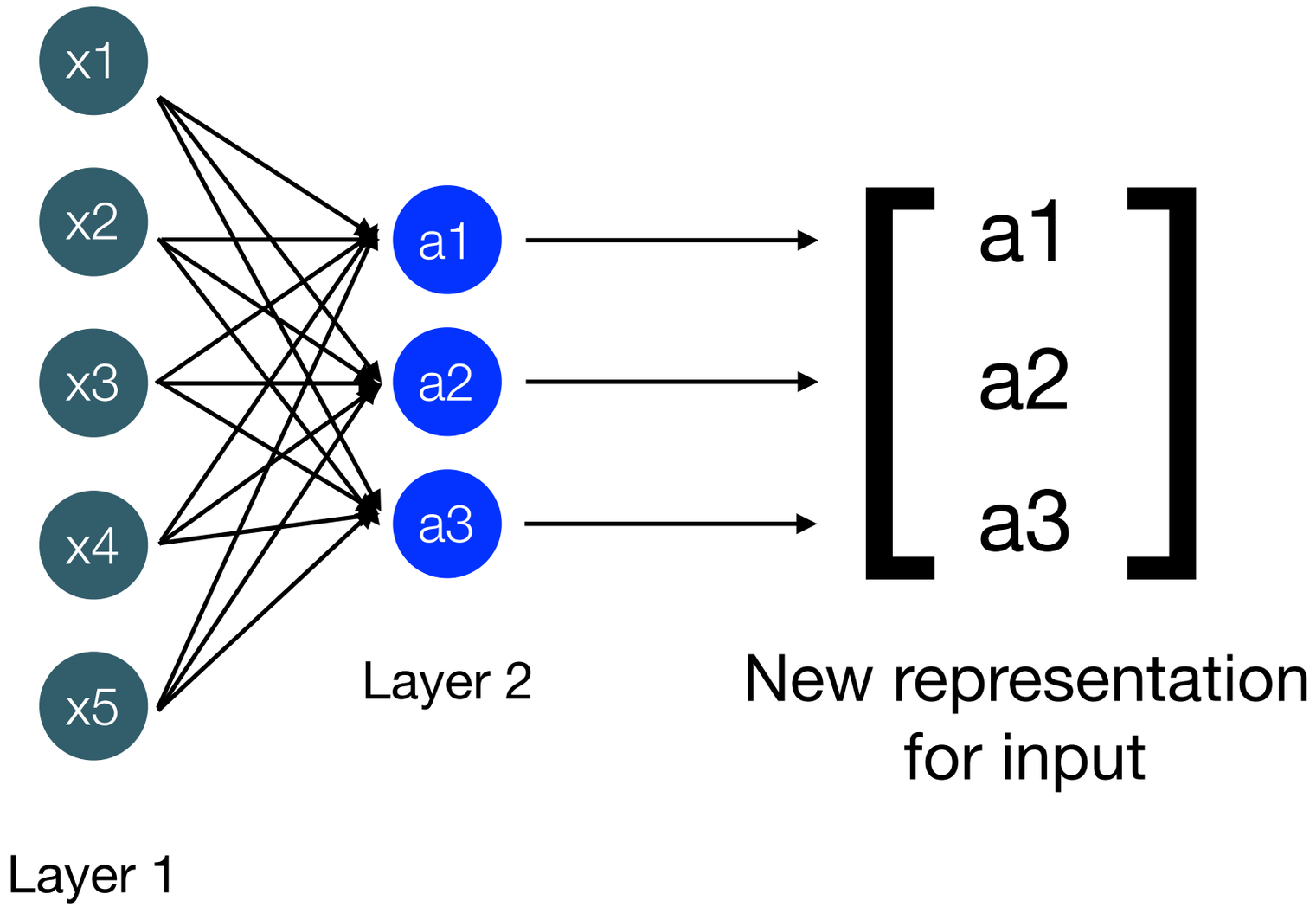
Deep Autoencoders



- Network is trained to output the input i.e. learn the identity functions.

- Constrain number of units in hidden layer, thus learning compressed representation.

Deep Autoencoders



Reference:
Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." Science 313.5786 (2006): 504-507.

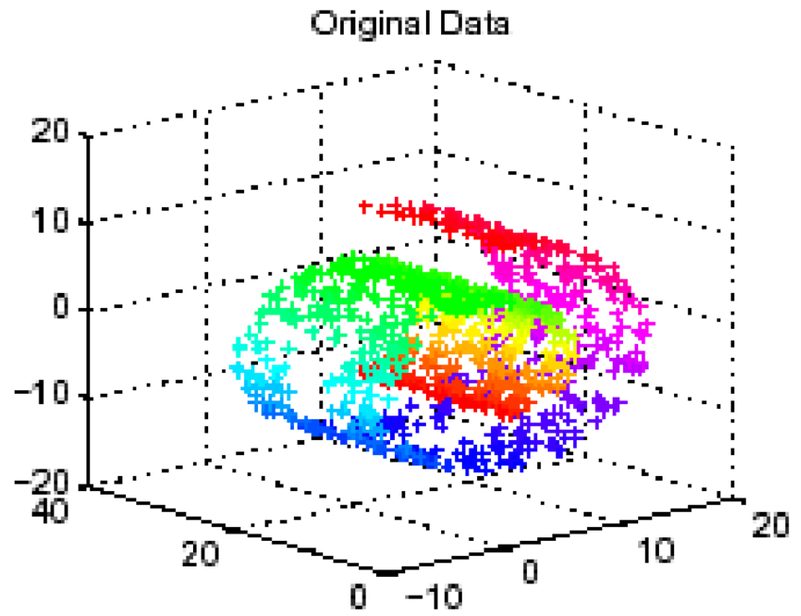
Feature Clustering

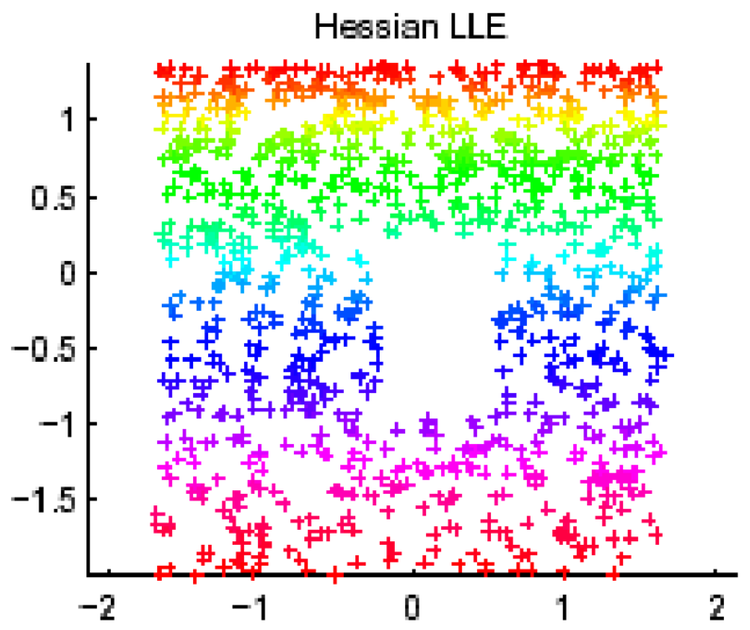
- Partition features into different clusters
- Features in the same cluster contain similar structural information of the given instances
- Obtained feature subset consists of features from variant clusters, so similarity between selected features will be low

Reference:
Cheung, Yiu-ming, and Hong Jia. "Unsupervised Feature Selection with Feature Clustering." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on. Vol. 1. IEEE, 2012.

Hessian Linear Local Embedding

- A non linear dimensionality reduction method
- Embeds a set of points from high dimensional space to low dimensional space such that projected point should have the same neighbourhood as the original point





Reference:
Donoho, David L., and Carrie Grimes. "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data." Proceedings of the National Academy of Sciences 100.10 (2003): 5591-5596.

User Interface in TMVA

TMVA::DataLoader* newLoader = loader->VarTransform(option_string)	
Variance Threshold	“VT(threshold_value)”
Deep Autoencoders	“AE(network_layout)”
Feature Clustering	“FC(number_of_dimensions)”
Hessian LLE	“HLLE(number_of_dimensions, number_of_neighbours)”

Thank you.