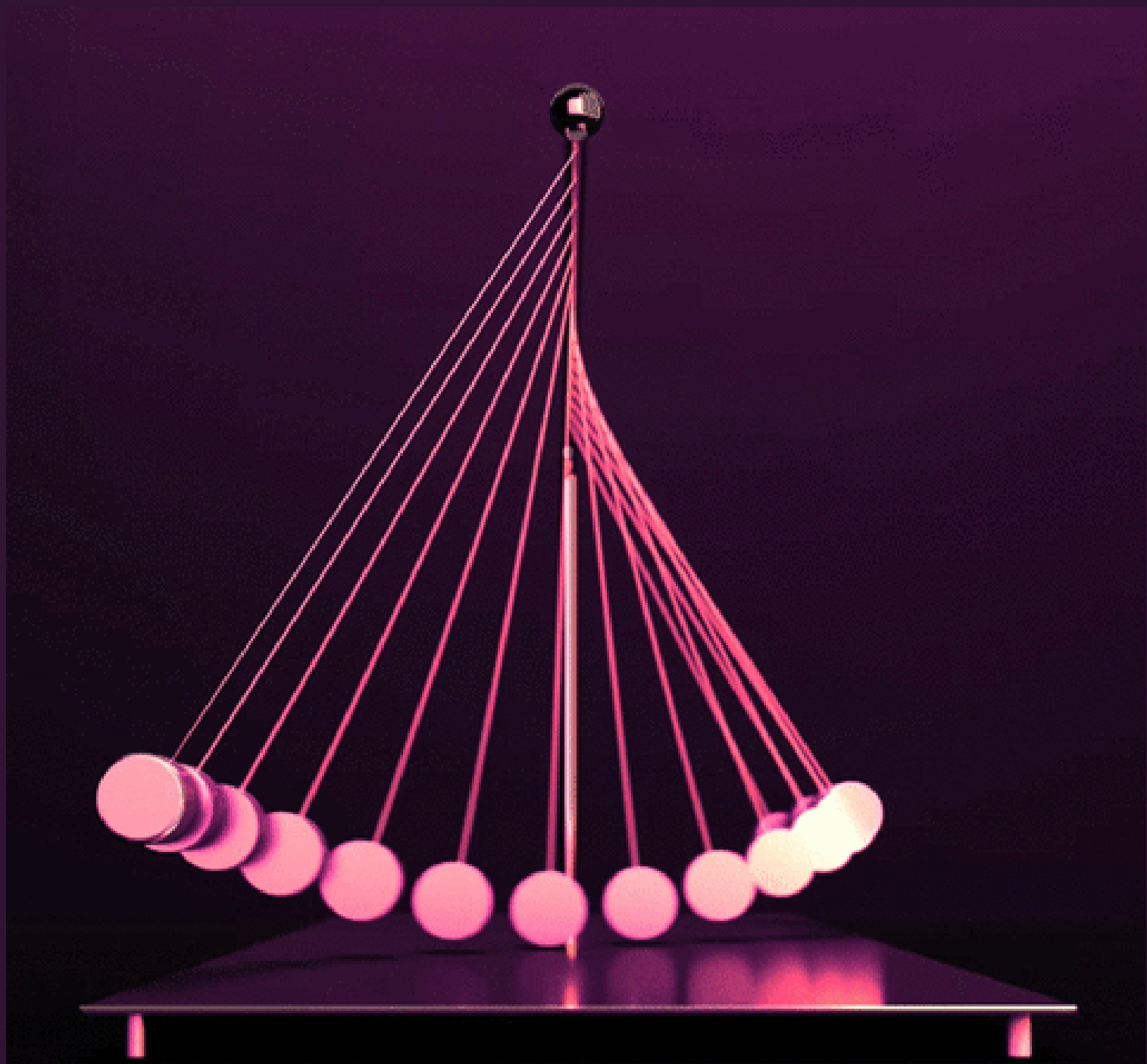
Machine learning Flaschcards - Pt 1



ShotNOTES

Gaussian Mixture Modeling

Fit through Expectation Maximization - basically assign labels then find mean of different labels.
Assumes data generated from a mixture of gaussians. Similar to K-means, but with information about covariance structure of data.
Can use BIC to estimate the optimal number of clusters.

Covariance can be: spherical - same amount of variance in each direction (K-means), diagonal - different features have different amounts of variance across clusters (only constraint is all variance assumed independent), full - no constraints (so can have correlated variance), tied - all clusters have same variance.

Variational Bayesian Gaussian Mixture

Maximizes lower bound on model evidence rather than data likelihood.
Uses EM algorithm, but adds regularization by using priors.
Concentration parameter controls how likely the model is to find more or less components (low value=few components).
Components modeled as from a dirichlet process.
Slower than GMM, but lets the model have more input in how many components to select.

Isomap

(Isometric Mapping). Form of manifold learning. Extension of Multi-Dimensional Scaling or Kernel PCA. Seeks a lower dimensional embedding that maintains geodesic distances between points. Complexity is >O(N**2)

Locally Linear Embedding

A series of local PCAs which are combined for a global non-linear embedding.
Complexity >O(N**2)
Regularization is a problem when the number of neighbors is greater than input dimensions - rank deficiency. Modified LLE designed to combat this.