**Markov Models**

Created: 10/04/2023 by Tom Lever

Updated: 10/04/2023 by Tom Lever

*Machine Learning*

Designing algorithms for inferring what is unknown based on knowns

Blend of Statistics and AI

Used in recognizing spam, recognizing handwriting, self-driving cars, blurring faces in Google Street View imagery, recognizing speech, recommending, and interpolation of climate data

Supervised vs. Unsupervised Learning

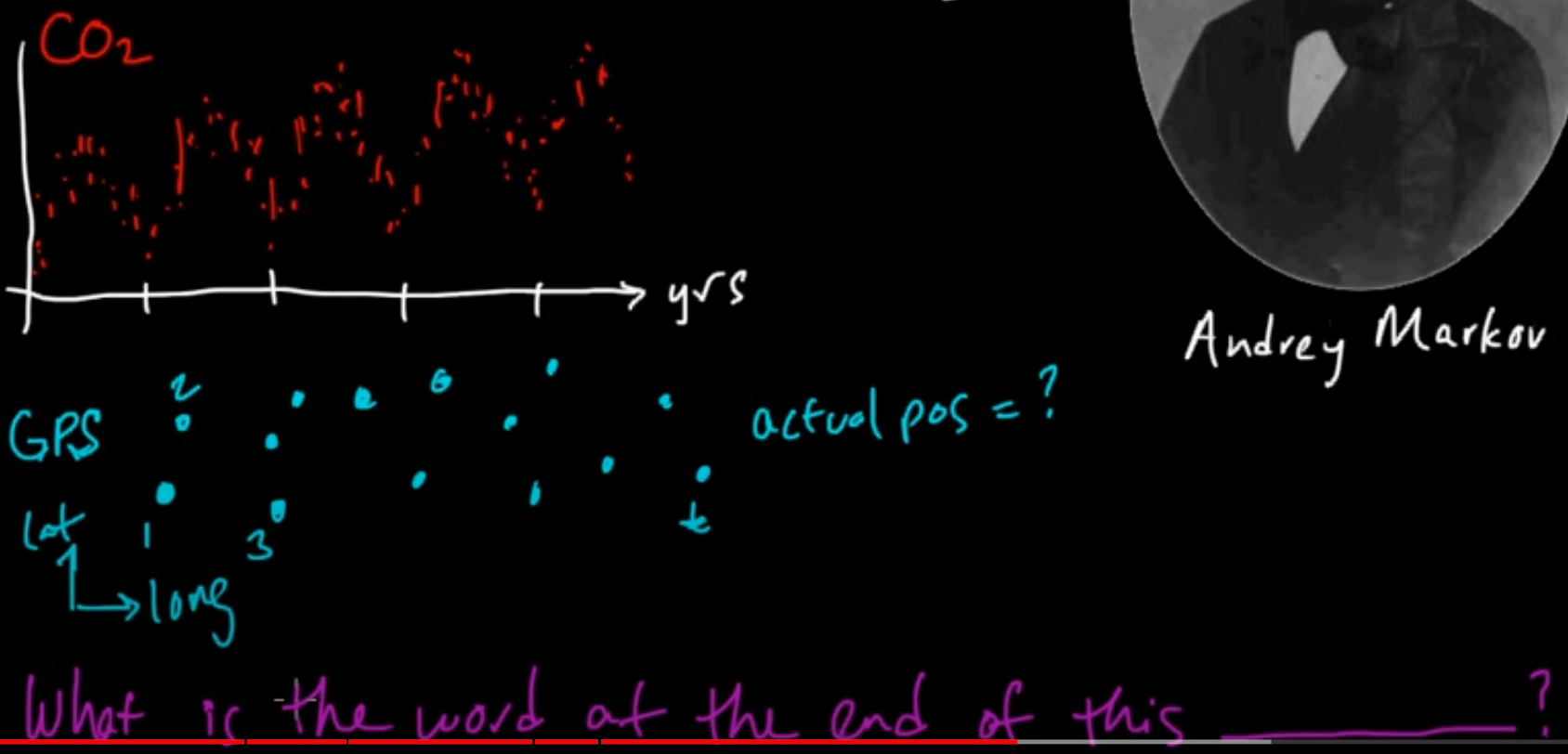
Supervised learning is a problem in which we choose a function based on a sequence of coordinate pairs where is a data point and is a target value that can be used to approximate a target value based on a data point .

Types of supervised learning include classification and regression. For classification, belongs to some finite set. For regression, .

The future is independent of the past, given the present.

Markov models deal with temporal data related to subjects like weather, finance, language, music.

Consider carbon-dioxide concentration in atmosphere vs. year, position over time, and completing sentences.



Consider tuple of sequential data . We model this data using random variables in . These random variables are not independent and identically distributed: carbon-dioxide concentration a little after a time is close to the carbon dioxide concentration at that time. The most accurate prediction of what will happen in the near future is what is happening now; the most accurate prediction of what is happening now is what happened in the recent past. Suppose depends on where are instants in time separated by a certain unit of time.

We assume that data points in occur at discrete instants in time. We assume that the possible values of the data points in comprise a finite set. Discrete random variables in form a first-order discrete-time Markov chain if the joint distribution of the random variables respect



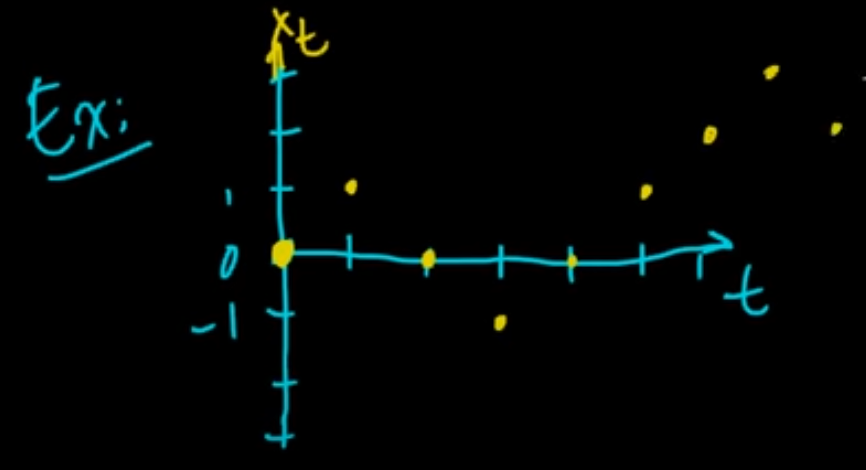
or the joint distribution factors as

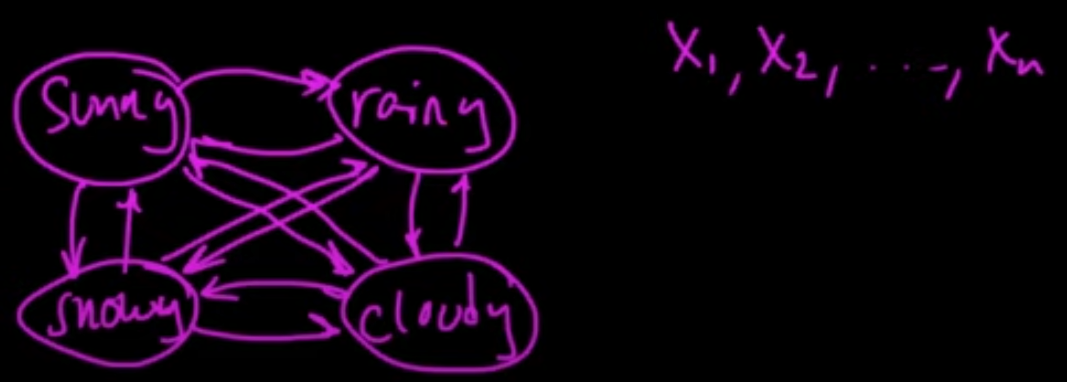
or

or is conditionally independent of given .

For example, is conditionally independent of given .

First-order discrete-time, discrete-space Markov chain (e.g., random walk on integers, weather)

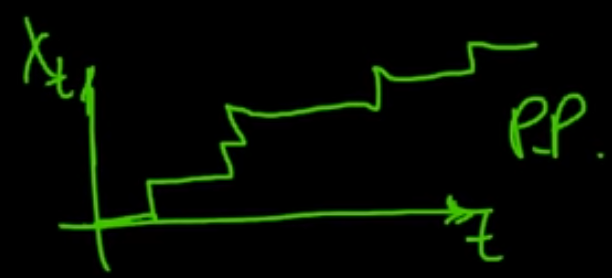




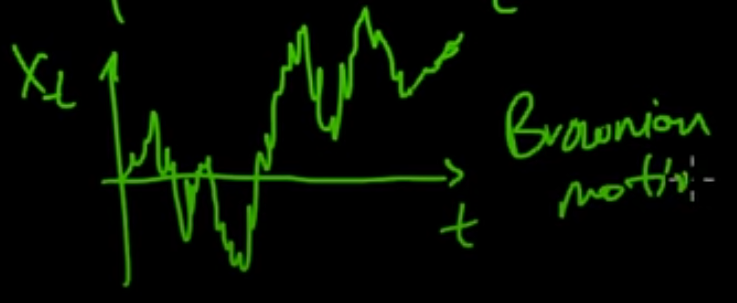
Second-order discrete-time, discrete-space Markov chain



First-order continuous-time, discrete-space Markov process (e.g., a Poisson process)



First-order discrete-time, continuous-space / state-space Markov process (e.g., 1D Brownian motion)

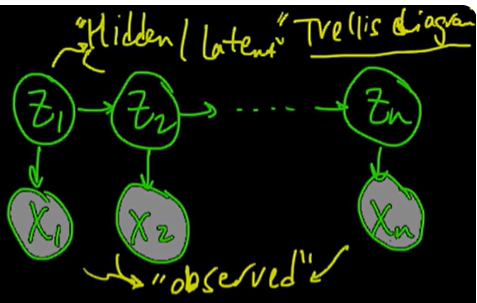


In our first graph above, the nodes represent the state of the system as it evolves. We assumed that our data is a subset of the true state of the system as it evolved. However, we can’t expect to perfectly observe the complete true state of the system. We collect noisy observations at instants of time that different from the true values of the system at those times. We can theorize that there are hidden / latent information / variables; the state of the system consists of an observed state and a hidden state. We can construct a Hidden Markov Model.

Consider a tuple of observed random variables . Each random variable takes on a discrete value, a real value, a vector of real values, etc. Data is a tuple of observed values.

Consider a tuple of hidden random variables . Each random variable takes on a discrete value in . Each value of a hidden variable occurs at a discrete time.

These variables / the joint distribution of these variables respects a trellis diagram / graphical model for a Hidden Markov Model:



Probability of transition from state to state .

Emission probability

Emission probability

is a probability density distribution of a set of real values, vector of real values, etc.

is a probability mass distribution of a set of discrete values

Initial probability mass distribution