Jan 21, 2020

CS 6241: Numerical methods for data science

Numerical methods:

- algo that we put on a computer to "solve a problem"
- usually continuous math, interplay with discrete

Data science

- ??
- analysis of some info that has already been collected for new insight

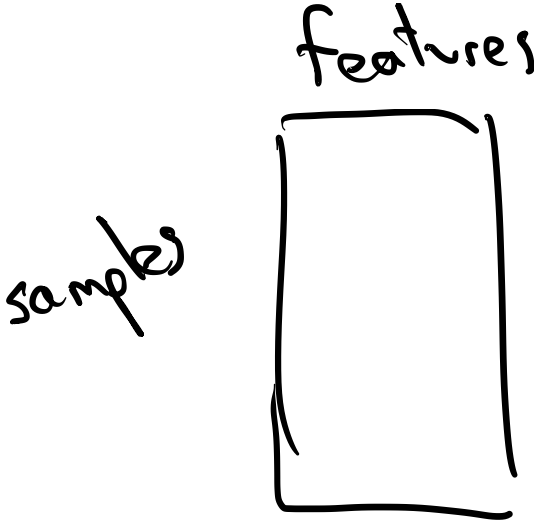
This class

(1) matrix methods — basic ML, opt
                    — dimensionality redux

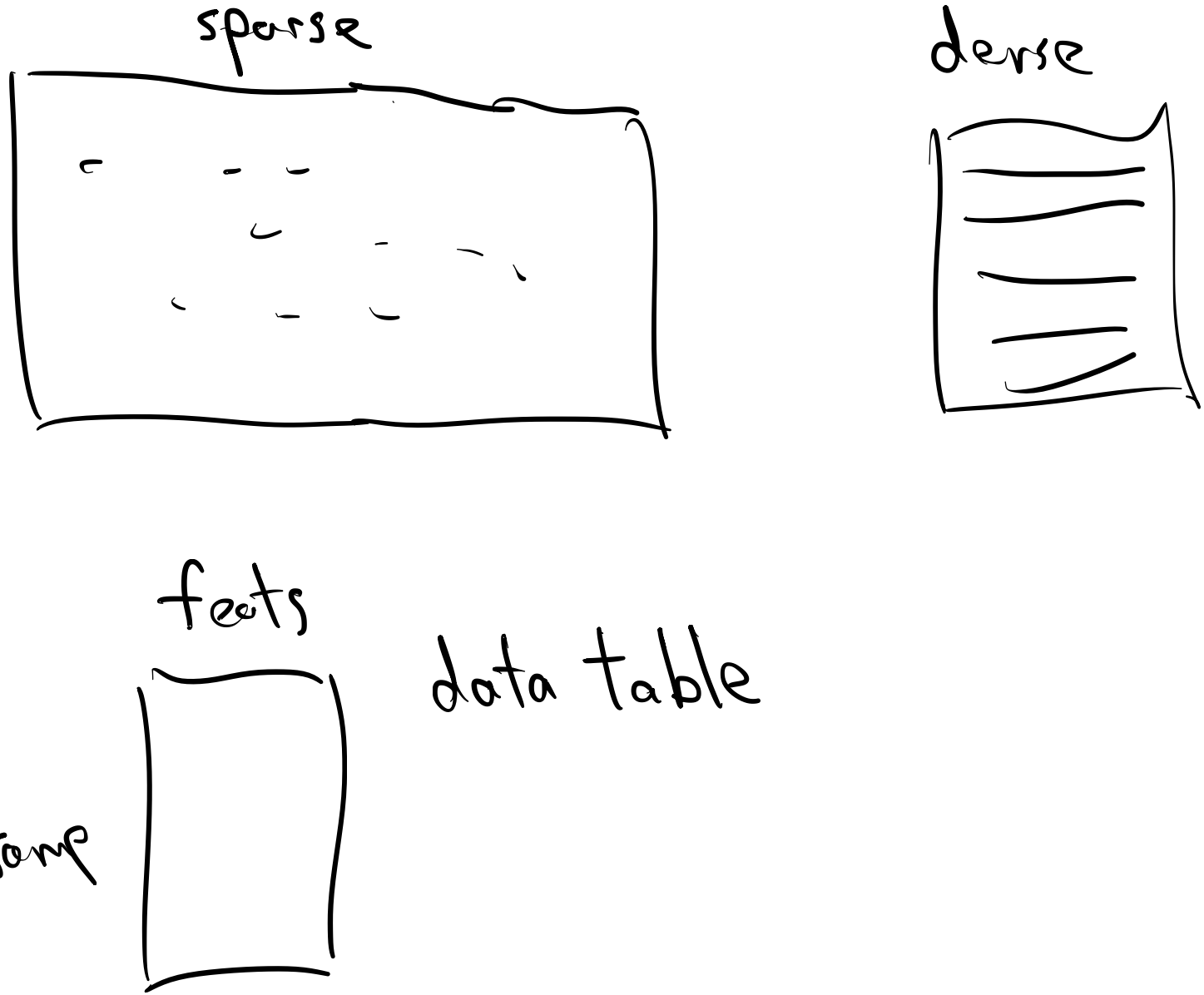
(2) network analysis — ranking
                     — clustering
                     — ML

special topics: GPs / kernel, time series, discrete choice

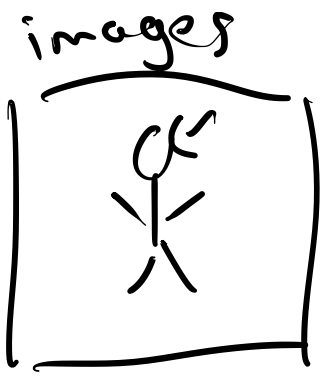
Two important objects: (1) Matrix
(2) Graph

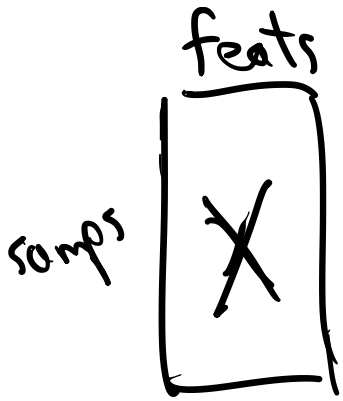


samples	feats	problem
animals in Vet lab	height, weight, food	sick?
patients	prescriptions	cluster new Rx



matrices with two-dim structure





$$\bar{X} = \left(I - \frac{1}{n} 1 1^T\right) X$$

$$S = \frac{1}{n-1} \bar{X}^T \bar{X}$$

$$\bar{X} = \left(I - \frac{1}{n} 1 1^T\right) X$$