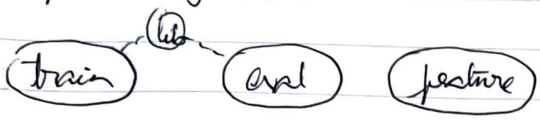
03/20/2024

Meeting Notes:

Refactoring Code



- convert svs to feature
- arg parsing
- parse annotations
- stream svs files — framing and rendering
- Alg: dct
- write to csv
- Geometry
- pixel to RGBA
- RGBA to feats via DCT
- parameter file parsing

---

train

- arg parsing
- parameter parsing
- load csv files
- ML training
- model I/O

eval

- arg parsing
- csv reader
- model loading
- ML classification
- Hypothesis writing
- Scoring
- Postprocessing

The way we cluster these libraries/function will decide what libraries we have.

Libraries
- arg parsing (rdkit_cmdline_tools)
- Model I/O (sklearn)
- ML alg (sklearn)
- SVS file I/O (simsage_tools)
- CSV write (pandas)
- Geometry (shapely)
- param fit (NEDC)
- annotation (ann_tools)

1) gen-feats -p param_file ~~*.list~~ svs.list ann.list
   → svs to csv

2) train -p param -m model svs.list ann.list

3) eval -p param -m model svs.list ann.list

4) score gdf.list hyp.list

Num. Th (1) (2) (3) (4)