Systematic comparison of soft vs. hard multi-task models for chromatin accessibility
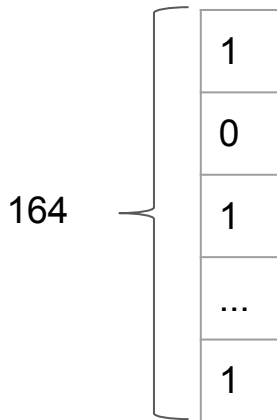
Ana Istrate, Michael Painter, Sebastian Le Bras

Dataset

- The same as the one from Bassett paper
- ~2,000,000 samples
  - 600 bp long
  - Label: binary vector that the 600 bp sequence has statistically significant chromatin accessibility in one of 164 cell types
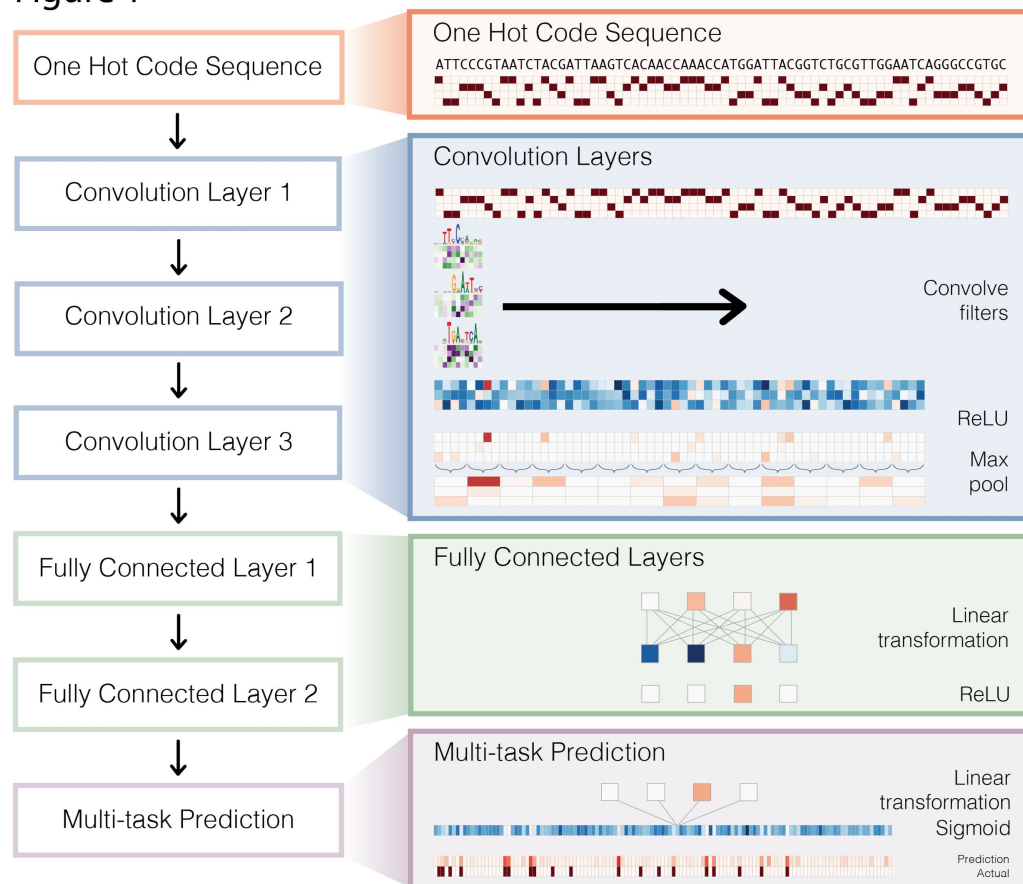
One Hot Code Sequence

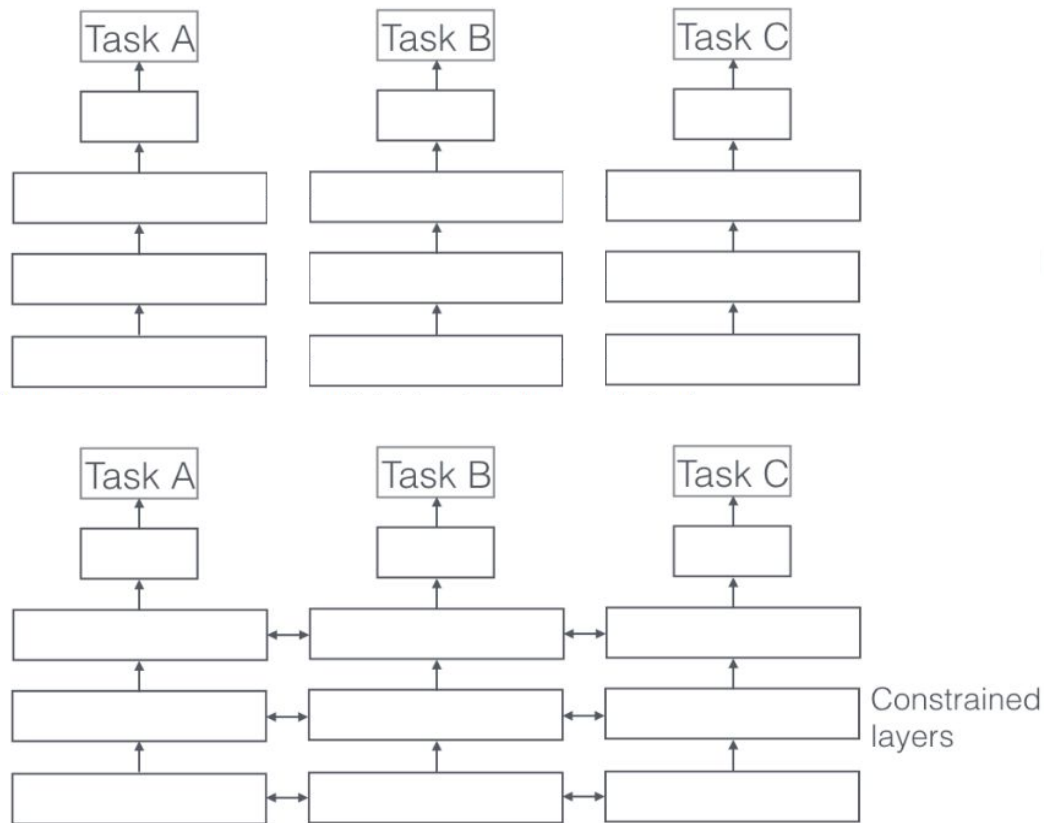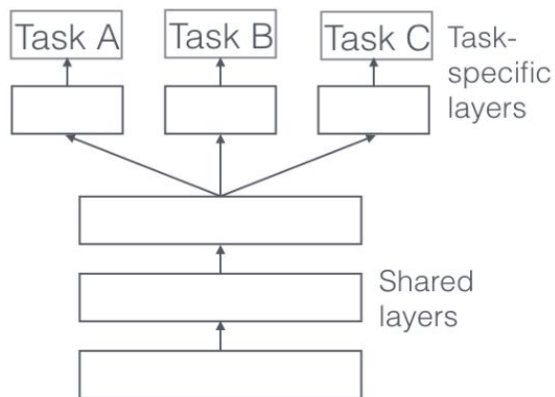ATTCCCGTAATCTACGATTAAGTCACAACCAAACCATGGATTACGGTCTGCGTTGGAATCAGGGCCGTGC

164 {
| 1 |
| 0 |
| 1 |
| ... |
| 1 |

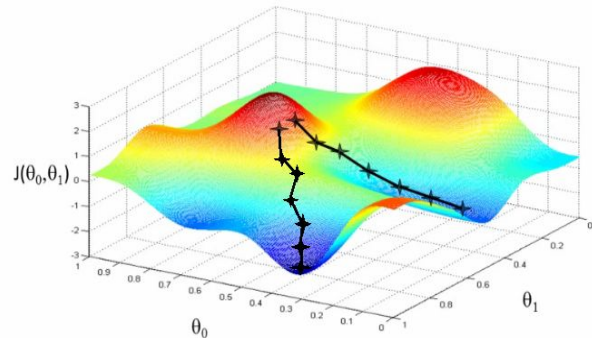# Basset Architecture

Figure 1

Architectures to Compare

- We will compare:
  - Single task models
  - Soft multi-task models
  - Hard multi-task models

Related work

- A Hard multi-task and a Single-task models exist.
- Single-task model outperforms the Hard multi-task model, because of the way the loss functions get trapped in local minima.

Goals of the project

1. Improve the Hard multi-task model to the level of the Single-task model
- Better starting points, different training techniques (train for one task at a time in each epoch, experiment with warm start on a task)
2. Build a soft-multi task model and compare it to the other two

Comparisons and Analysis

- Accuracy/Training Loss/AUC
- Training time complexity
- Interpretations of hidden layers learned (how similar are they between models)
- Where do the gradients 'die'?
  - Is it when gradients for *all* subtasks are near zero?

Questions?