- Draft -

Regularization

* Aims to solve the problem of Catastrophic Forgetting
* Some attempts on tackling plasticity [Lewandowski / Kumar]
* Two general approaches
  + Parameter regularization
    - Penalty term on the estimated parameters theta
    - Often depends on available data
  + Function space regularization
* General assumption is overparameterization
  + Guarantees a solution for Loss

Parameter Regularization

* Multiple adaptations of L2 (Ridge-Penalty) have been developed [Kirkpatrick, Kumar, …]
  + Centring around the previous tasks’ parameters [Zhao] or around an initialized set of parameters [Kumar]
  + Use of weight matrix H that is responsible setting the “importance” of each parameter [Zhao]
  + Special case is minimum norm [Lin] where y = X\*beta\_hat
  + Special case is continual ridge regression with equal hyperparameter lambda for all tasks