Stochastic Frank-Wolfe for DNN training

Constraints

The paper lists several options for constraints (feasible regions for parameters). While this is by no means a complete list, it does cover several common use-cases. How to best solve the linear optimization problem in the algorithm depends on the contraint type.

1. **Lp-norm ball**:
   Description:
   convex region by a fixed/ bounded Lp norm of the parameters.

   Expected result:
   p=1 -> sparse weights with many being exactly 0.
   p=2 -> many weights close to 0.
   p=∞ -> Hypercube constrains maximum value of each weight. This supposedly helps to prevent overfitting.

   Notes:
   ---

2. **K-sparse polytope**:
   Description:
   Convex hull of intersection of the L1 ball and a hypercube. Spanned by all vectors with exactly K nonzero entries.

   Expected result:
   Exactly K nonzero weights.

   Notes:
   K is a hyperparameter that needs to be set before training.

3. **K-norm ball**:
   Description:
   Convex hull of union of the L1 ball and a hypercube.

   Expected result:
   Combination of properties of L1 norm and hypercube: sparsity with constrained magnitude of weights.

   Notes:
   ---

4. **Unit simplex/ probability simplex**:
   Description:
   n/ n-1 dimensional simplex

   Expected result:
   Sum of weights is 1, weights represent probabilities.

Notes:

---

5. **Permutahedron**:

   Description:
   Polytope spanned by all permutations of the coordinates of the vector $(1, 2, ..., n)$.

   Expected result:

   Notes:
