

Your Classifier is Secretly an Energy Based Model and You Should Treat it Like One

1. Как задается распределение энергетической модели? Чем отличается подход авторов Your Classifier is Secretly an Energy Based Model от стандартного подхода к определению энергетической модели?
2. Выведите формулу для градиента плотности энергетической модели. Как он приближается на практике?
3. Какие преимущества по сравнению со стандартными моделями классификации обнаружили у своей модели авторы Your Classifier is Secretly an Energy Based Model?
4. В чем заключаются основные минусы энергетических моделей?

Gradient Estimation with Stochastic Softmax Tricks

1. Пусть E_1, \dots, E_n независимые случайные величины с экспоненциальным распределением с параметрами $\lambda_1, \dots, \lambda_n$ соответственно. Как будут распределены случайные величины $V = \min_i E_i$ и $i^* = \arg\min_i E_i$?
2. Предположим, что функция потерь имеет вид $E_{\{p(z | \theta)\}} f(z)$. Какие подходы возможны при оценке её градиента? Опишите также ограничения этих подходов.
3. В чем заключается Gumbel Softmax Trick? Какая связь между Gumbel Softmax Trick и Stochastic Softmax Trick?

Learning with Differentiable Perturbed Optimizers

1. Зачем авторы статьи Learning with Differentiable Perturbed Optimizers используют возмущенный максимизатор? Какие задачи позволяет решить предложенная ими техника?
2. Выпишите формулу функции потерь Фенхеля-Янга и объясните все ее обозначения. Как авторы предлагают вычислять градиент функции потерь Фенхеля-Янга на практике?
3. Опишите любой из 3 экспериментов из работы Learning with Differentiable Perturbed Optimizers и выводы, к которым приводят авторы.

Generative Pretraining from Pixels

1. Как авторы Generative Pretraining from Pixels предлагают предобрабатывать изображение перед применением трансформера? В чем заключается мотивация этих преобразований?
2. Какие два подхода авторы рассматривают для предобучения трансформеров на изображениях? Как авторы строят представление изображения по выходам трансформера? Как два рассмотренных подхода показывают себя в экспериментах?
3. Как в работе Generative Pretraining from Pixels осуществляется дообучение модели?