

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

1. Как авторы Your Classifier is Secretly an Energy Based Model определяют совместную плотность $p_{\theta}(x, y)$? Как получить (с точностью до нормировочной константы) $p_{\theta}(x)$ и $p_{\theta}(y | x)$? Какое выражение градиента логарифма плотности $p_{\theta}(x)$ авторы используют при обучении?
2. Опишите алгоритм обучения модели из работы Your Classifier is Secretly an Energy Based Model. Какую целевую функцию он оптимизирует?
3. Какие трудности возникают при обучении модели из работы Your Classifier is Secretly an Energy Based Model?

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. Выпишите альтернативное определение функции Softmax, используемое авторами Stochastic Softmax Tricks для обобщения Gumbel 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. Какие свойства возмущенных максимизаторов описывают авторы Learning with Differentiable Perturbed Optimizers?

Generative Modeling by Estimating Gradients of the Data Distribution

1. Какие нововведения предлагают авторы Generative Modeling by Estimating Gradients of the Data Distribution в своем подходе к задаче генерации?
2. Какие основные проблемы возникают при обучении генеративных моделей, приближающий градиент плотности данных $\nabla_x \log p(x)$?
3. Как добавление гауссовского шума влияет на обучение в работе Generative Modeling by Estimating Gradients of the Data Distribution?
4. Кратко опишите алгоритм с динамикой отжига Ланжевена для генерации данных из обученной авторами модели NCSN.