

Вариационные Автоенкодеры с мультиномиальным Prior'ом.

Денис Мазур

Ноябрь 2018

Аннотация

Метод вариационных автоэнкодеров является, вероятно, самым эффективным среди методов глубинного обучения без учителя...

Содержание

1 Вступление	1
2 Обучение нейронных сетей	1
2.1 Градиентный метод	2
2.2 Градиентный метод в обучении нейронных сетей	2
3 Автоенкодеры	3
3.1 Почему это работает?	3
4 Вариационные автоенкодеры	3

1 Вступление

Так как прочтение данной работы не предполагает ознакомленность с методами машинного обучения, будет дан поверхностный разбор некоторых ключевых принципов и методов, необходимых для понимания вариационных автоенкодеров.

От читателя предполагается понимание математического анализа, линейной алгебры и теории вероятности.

2 Обучение нейронных сетей

Для базового понимания автоенкодеров не требуется понимание принципа работы нейронных сетей, по этой причине объяснение устройства нейронных сетей было решено опустить и уделить больше внимания принципу их

обучения. Для интересующихся, литература на тему нейронных сетей будет приложена в библиографии.

Для простоты устройсто нейронной сети можно редуцировать до некоторой функции, имеющей параметры W , принимающей значения X и выдающей значения \hat{X} . Физический смысл X - данные, подаваемые на вход сети, \hat{X} - выход нейронной сети. Каждому элементу $x \in X$ соответствует элемент $y \in Y$, и задачей обучения нейронной сети является подобрать параметры W нейронной сети так, чтобы они минимально или не отличались от Y . Также можно сформулировать задачу следующим образом:

$$d(f(x_n), y_n) \rightarrow \min$$

Также, можно поставить задачу как поиск оптимальных параметров W :

$$\underset{W}{\operatorname{argmin}} d(f(x_n), y_n)$$

Где $f(x)$ - нейронная сеть, а $d(x, y)$ определенная нами функция расстояния (отличия) между выходом нейронной сети и настоящим ответом, называемая функцией потерь. Важно понимать следующие свойства функций $f(x)$ и $d(x, y)$, они обе дифференцируемые и определены на всех X , а это значит, что мы к параметрам нейронной сети W можем применять любые методы оптимизации функций, например, градиентный метод.

2.1 Градиентный метод

Градиентный метод является методом решения задач вида $\min_{a \in \mathbb{R}^n} g(x)$. Алгоритм градиентного метода начинается с выбора параметров a , это может делать как и из какого либо представления того какими примерно эти параметры должны быть, так и случайным образом. Далее идет так называемый "шаг" градиентного метода. Из параметров a вычитается значение градиента функции $g(a)$ в точке текущего значения a :

$$a_{n+1} = a_n - \nabla g(a_n)$$

Чтобы не "перескочить" через минимум функции во время итерации метода, градиент $\nabla f(x)$ обычно умножается на какую-то константу α :

$$a_{n+1} = a_n - \alpha \nabla g(a_n)$$

2.2 Градиентный метод в обучении нейронных сетей

Теперь о том как градиентный метод может применяться в обучении нейронных сетей.

Функцию потерь нейронной сети $d(f(x_n), y_n)$ можно записать в виде, $d(f(x_n, W), y_n)$, до этого, для простоты, записи параметры нейронной сети W не записывались как аргумент, хотя, очевидно, им являются. Поясню,

что X и Y являются константными значениями, так как являются выборкой наших данных и известны нам заранее. Таким образом, единственный аргумент, который мы можем изменять, минимизируя функцию потерь - W .

На практике, оптимизировать функцию потерь градиентным спуском "в лоб" не получится. У этого есть несколько причин, одна из главных - риск переобучения модели (ситуация при которой модель показывает идеальный результат на обучающей выборке, но плохой на валидации). Но помимо этого, выборка данных чаще всего слишком велика и не помещается в память компьютера, по этой причине на каждой итерации градиентного метода выбирается случайная подвыборка тренировочных данных и над ней совершается шаг градиентного метода. Данный метод называется стохастическим градиентным спуском.

3 Автоенкодеры

Автоенкодеры являются архитектурой нейронных сетей, использующейся преимущественно для кодирования многомерных данных в данные меньшей размерности. Автоенкодеры состоят из двух отдельных нейросетей, первая - энкодер, преобразующая исходные данные в сжатую репрезентацию себя. Обозначим энкодер как $e(x) \rightarrow z$. Вторая - декодер, преобразующая сжатую репрезентацию данных z в их первоначальную форму. Обозначим декодер как $d(z) \rightarrow \hat{x}$.

$$e(x) = z$$

$$\|z\| < \|x\|$$

$$d(e(x)) \approx x$$

3.1 Почему это работает?

Архитектура автоенкодеров исходит из того, что изначальные данные содержат больше информации, чем необходимо для их кодирования, а значит при их кодировании мы не теряем данные. Это становится интуитивно понятно, если представить себе набор данных из изображений геометрических фигур. Так как для их кодирования нам надо знать лишь вид фигуры, цвет и положение.

4 Вариационные автоенкодеры

Есть два подхода к пониманию вариационных автоенкодеров. Один несколько более инженерный и практичный, другое более философски-идеоматический с точки зрения статистики и Байесовского машинного обучения. Эта работа сфокусируется на втором по той причине, что он дает большее понимание

того, что мы хотим достичь нашей работой, а также дает более полное понимание их принципе работы.