

Анализ неструктурированных данных

Семинар 10

Вариационный автокодировщик

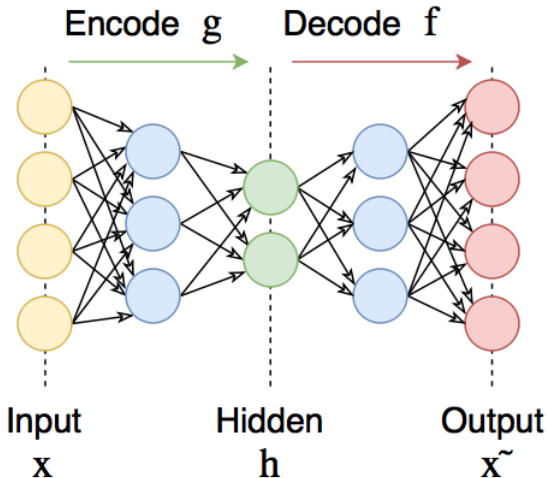
Мурат Апишев (mel-lain@yandex.ru)

НИУ ВШЭ (ГУ)

21-22 ноября, 2018

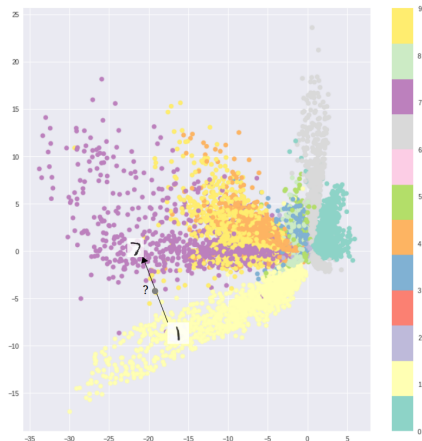
Автокодировщик

- ▶ Две связанных нейронных сети
- ▶ *Кодировщик* переводит объект в сжатое представление
- ▶ *Декодировщик* пытается из этого представления восстановить исходный объект



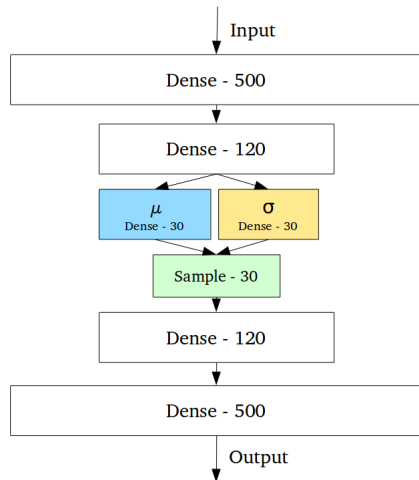
Автокодировщик как генеративная модель

- ▶ Нейросеть-декодировщик хочется использовать для генерации новых объектов из скрытого представления
- ▶ Обычный автокодировщик для этого подходит плохо:
 1. Построенное им представление не является непрерывным
 2. Кластеры, соответствующие разным классам объектов, разнесены в этом пространстве далеко друг от друга

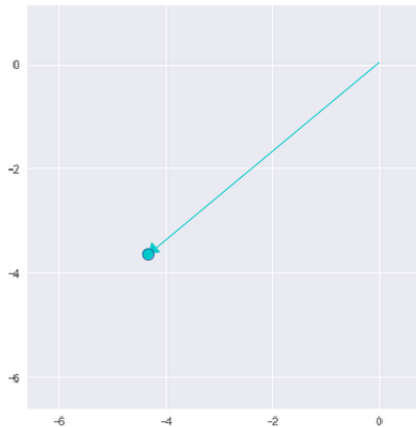


Вариационный автокодировщик

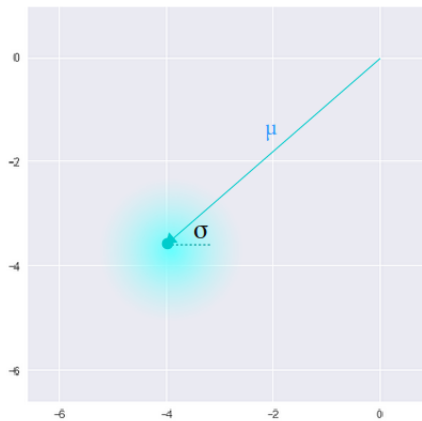
- ▶ Эти проблемы решает *вариационный автокодировщик*
- ▶ Его сеть-кодировщик не генерирует вектор представления
- ▶ Вместо этого он генерирует параметры его распределения: вектор средних и вектор стандартных отклонений
- ▶ Далее представление генерируется на основе этих параметров
- ▶ В результате получается непрерывное пространство представлений



AE vs. VAE



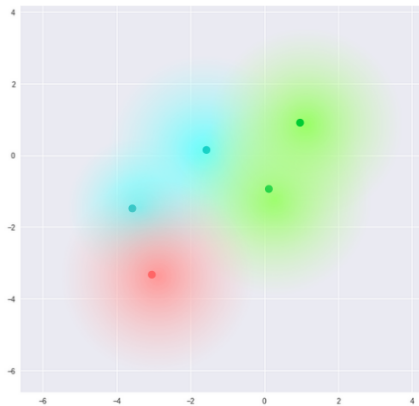
Standard Autoencoder
(direct encoding coordinates)



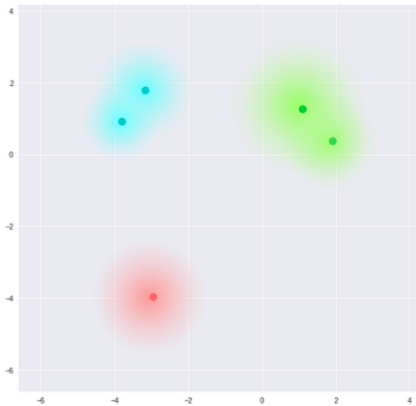
Variational Autoencoder
(μ and σ initialize a probability distribution)

Вид распределения

- ▶ Проблема непрерывности решилась
- ▶ Но кластеры всё ещё будут получаться отдалёнными



What we require



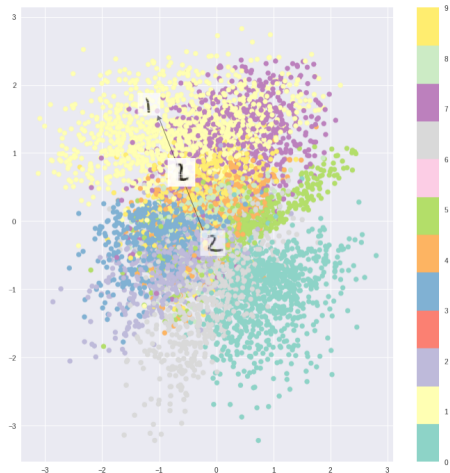
What we may inadvertently end up with

Оптимизируемый функционал

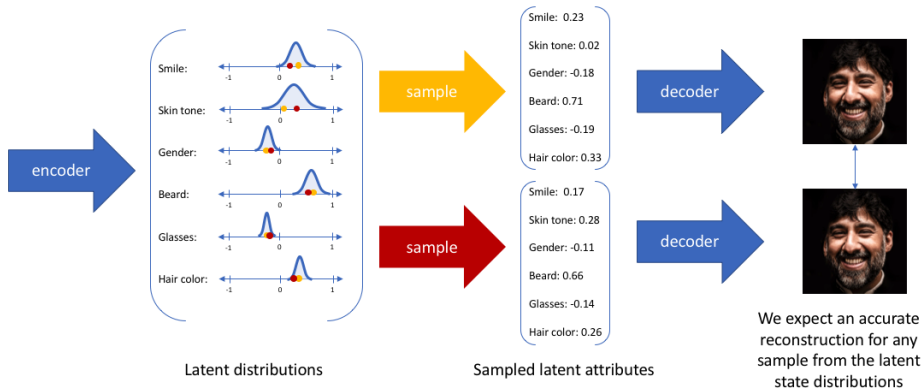
- ▶ Потребуем, чтобы векторы μ и σ были близки к параметрам стандартного нормального распределения
- ▶ Так же потребуем, чтобы сеть-декодер восстанавливала входные объекты
- ▶ Итоговый функционал качества:

$$\mathcal{L}(x) = \sum_i (x_i - \hat{x}_i)^2 + \sum_j \text{KL}(q_j(z|x) || \mathcal{N}(0, 1))$$

- ▶ i – индекс в размерности объекта,
 j – в размерности представления

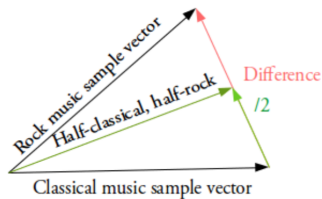


Зачем нужны непрерывность и близость

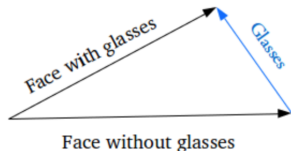


Зачем нужны непрерывность и близость

- ▶ Пусть мы кодируем музыку и хотим сгенерировать среднее между двумя жанрами
- ▶ Для этого достаточно прибавить к вектору μ одного жанра половину вектора разницы между ними
- ▶ Пусть мы кодируем фотографии лиц
- ▶ Разность векторов средних одинаковых фотографий с очками и без даст вектор, описывающий появление очков

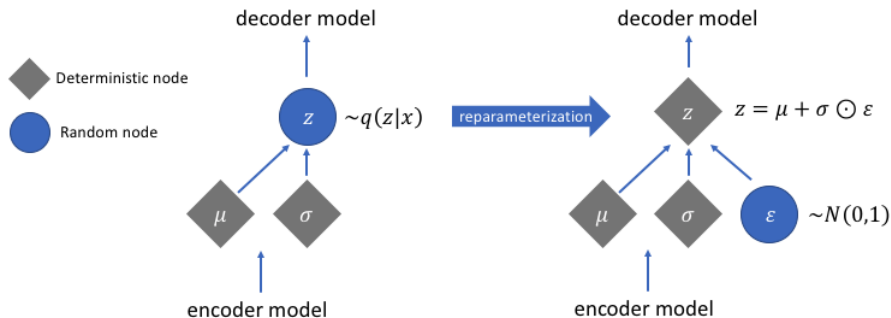


Interpolating between samples



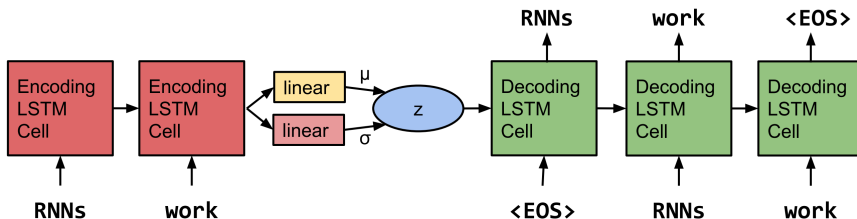
Трюк с репараметризацией

- ▶ Back-propagation не может пройти через стохастический элемент
- ▶ Идея: вынесем стохастичность в отдельную случайную величину
- ▶ Параметры станут детерминированными \Rightarrow дифференцируемыми



VAE и генерация текстов

- ▶ На вход кодировщику подадим векторы предложений, полученных с помощью LSTM
- ▶ Декодировщик на основе вектора представления будет генерировать предложение
- ▶ Ссылки на примеры реализаций: [1](#), [2](#)



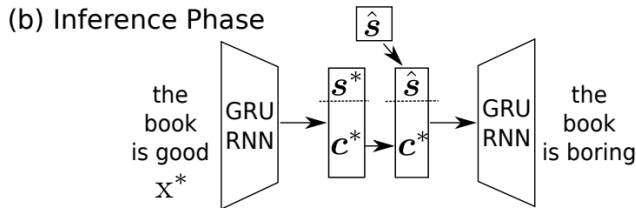
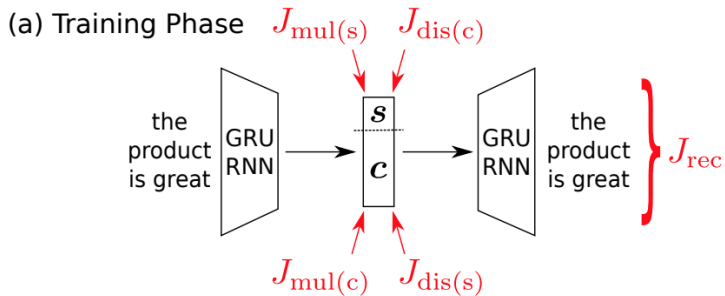
Text style transfer с помощью VAE

- ▶ К сожалению, на текущий момент VAE для текстов далеко не так хороши, как для картинок
- ▶ Даже в современных статьях «перенос стиля» ограничивается переносом тональности
- ▶ В рассматриваемой работе тональность бинарная
- ▶ Идея в том, чтобы формировать векторы стиля и контента по-отдельности
- ▶ Также хочется, чтобы каждое пространство содержало только информацию о кодируемой сущности (стили или контент)
- ▶ Для этого используется комбинация нескольких функций потерь

Функции потерь

- ▶ J_{rec} – восстановление входа
- ▶ $J_{\text{mul}(s)}$ – предсказание верной тональности
- ▶ $J_{\text{dis}(c)}$ – удаление информации о стиле из пространства для контента:
 1. Обучаем классификатор, распознающий стиль по вектору контента
 2. Обучаем кодировщик создавать такое представление для контента, чтобы классификатор не мог из него найти стиль
- ▶ $J_{\text{mul}(c)}$ – предсказание верного контента:
 - ▶ Под контентом понимается нормированный вектор «мешка слов» без стоп-слов и слов, специфичных для стиля (тональности)
 - ▶ Отличия от J_{rec} : восстанавливается только «мешок слов» (не всех) и контролируется только пространство представления контента
- ▶ $J_{\text{dis}(s)}$ – удаление информации о контенте для стиля (предсказывается тот же нормированный вектор «мешка слов»)

Архитектура сети



Примеры результатов

Original (Positive)
the food is excellent and the service is exceptional
the waitresses are friendly and helpful
the restaurant itself is romantic and quiet
great deal
both times i have eaten the lunch buffet and it was outstanding
Original (Negative)
the desserts were very bland
it was a bed of lettuce and spinach with some italian meats and cheeses
the people behind the counter were not friendly whatsoever
the interior is old and generally falling apart
they are clueless

VAE Transferred (Negative)
the food was bland and i am not thrilled with this
the waitresses are rude and are lazy
the restaurant itself was dirty
no deal
both times i have eaten here the food was mediocre at best
VAE Transferred (Positive)
the desserts were very good
it was a huge assortment of flavors and italian food
the people behind the counter is friendly caring
the interior is old and noble
they are genuinely professionals