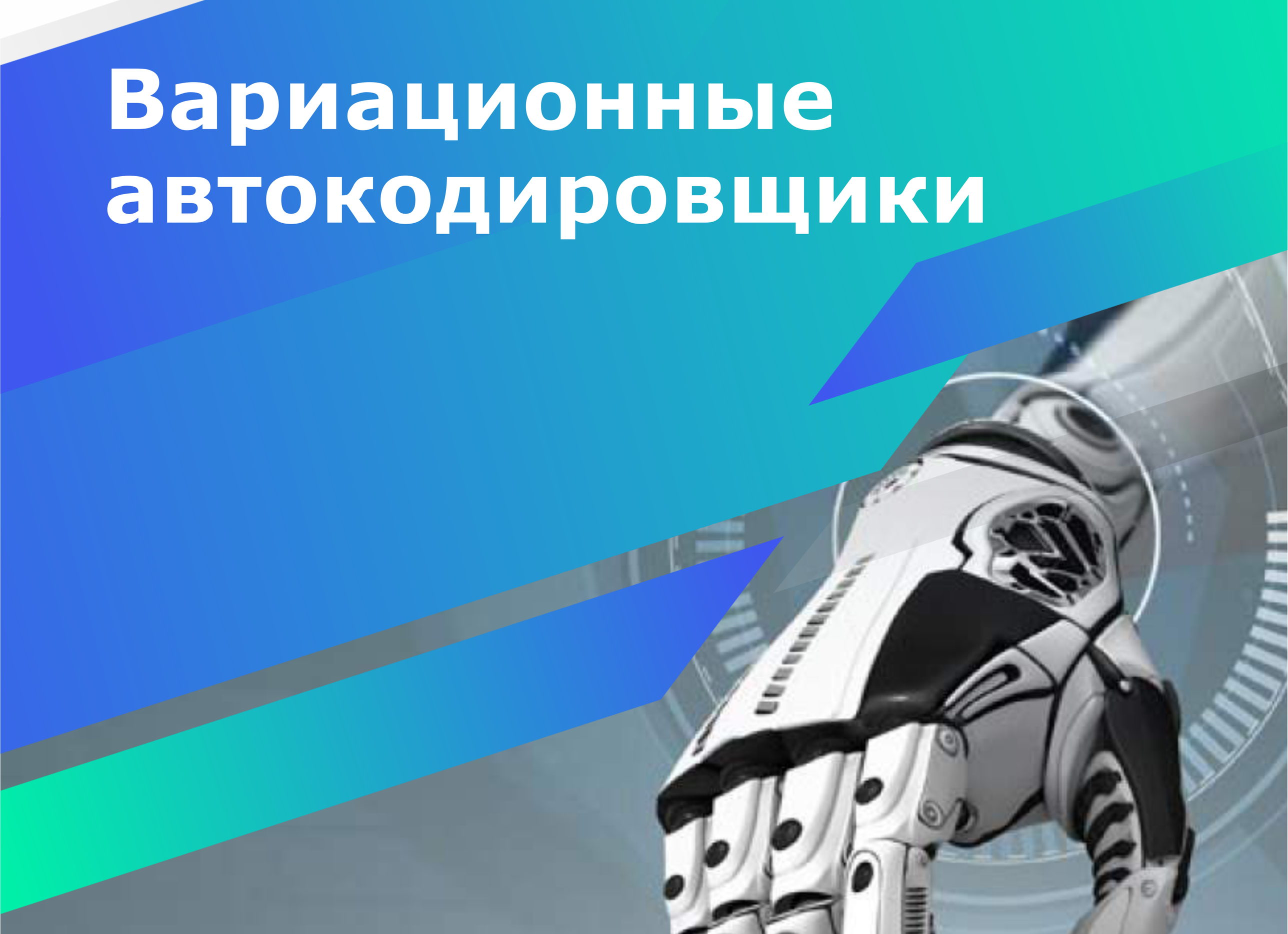
УНИВЕРСИТЕТ
искусственного
интеллекта

Вариационные автокодировщики

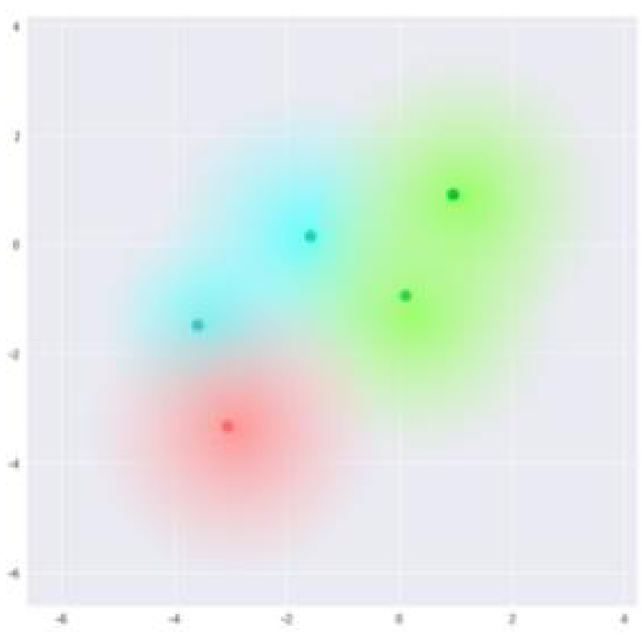


Вариационные автокодировщики

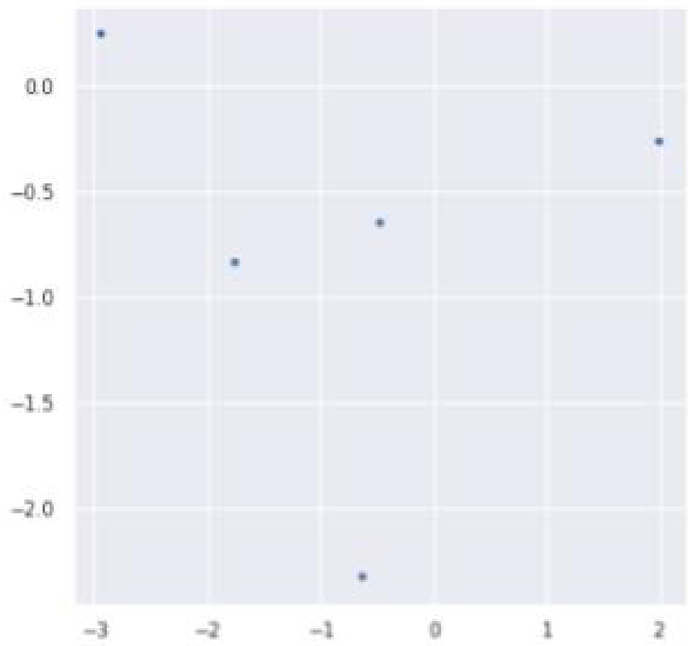
Вариационный автокодировщик (англ. *Variational Autoencoder*, *VAE*) — автокодировщик (генеративная модель, которая учится отображать объекты в заданное скрытое пространство и обратно), основанный на вариационном выводе.

Модель VAE находит применение во многих областях исследований: от генерации новых человеческих лиц до создания полностью искусственной музыки.

Вариационный автоэнкодер (VAE) имеет одно уникальное свойство, которое отличает его от стандартного автоэнкодера. Именно это свойство делает вариационные автоэнкодеры столь полезными при генерации данных: их **скрытое пространство** по построению **является непрерывным**, позволяя выполнять случайные преобразования и интерполяцию.



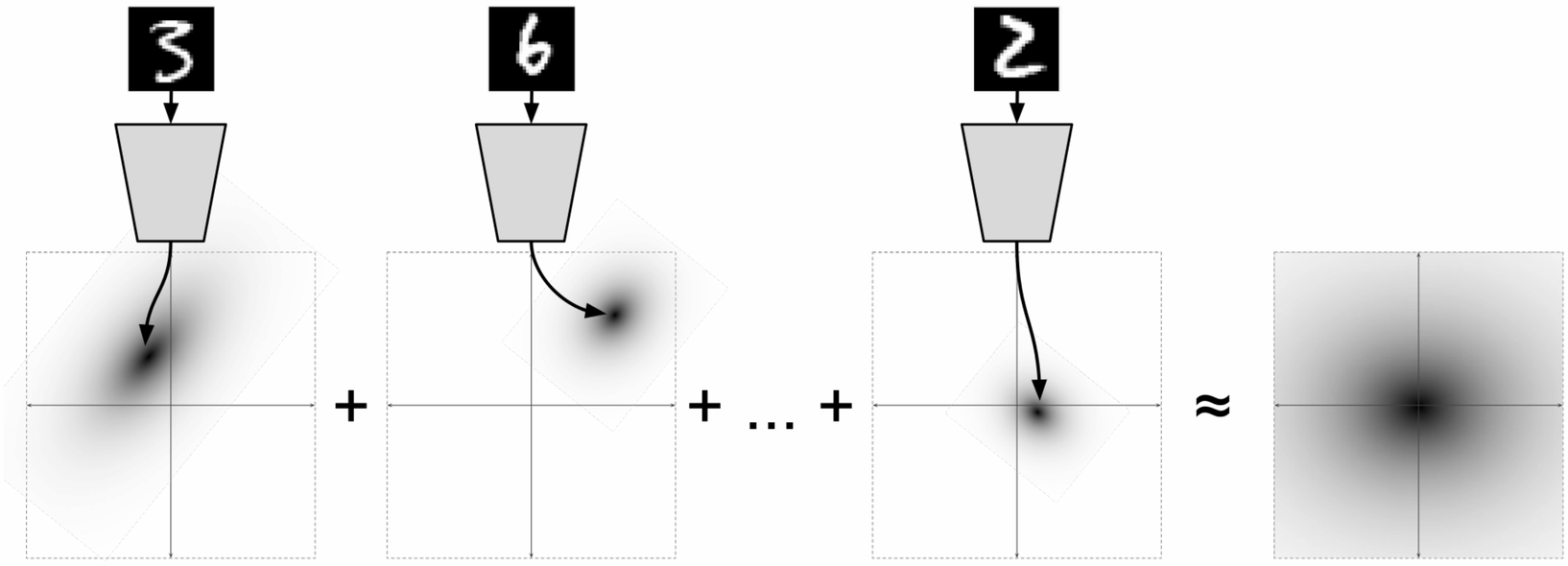
Расположение объектов в скрытом пространстве VAE



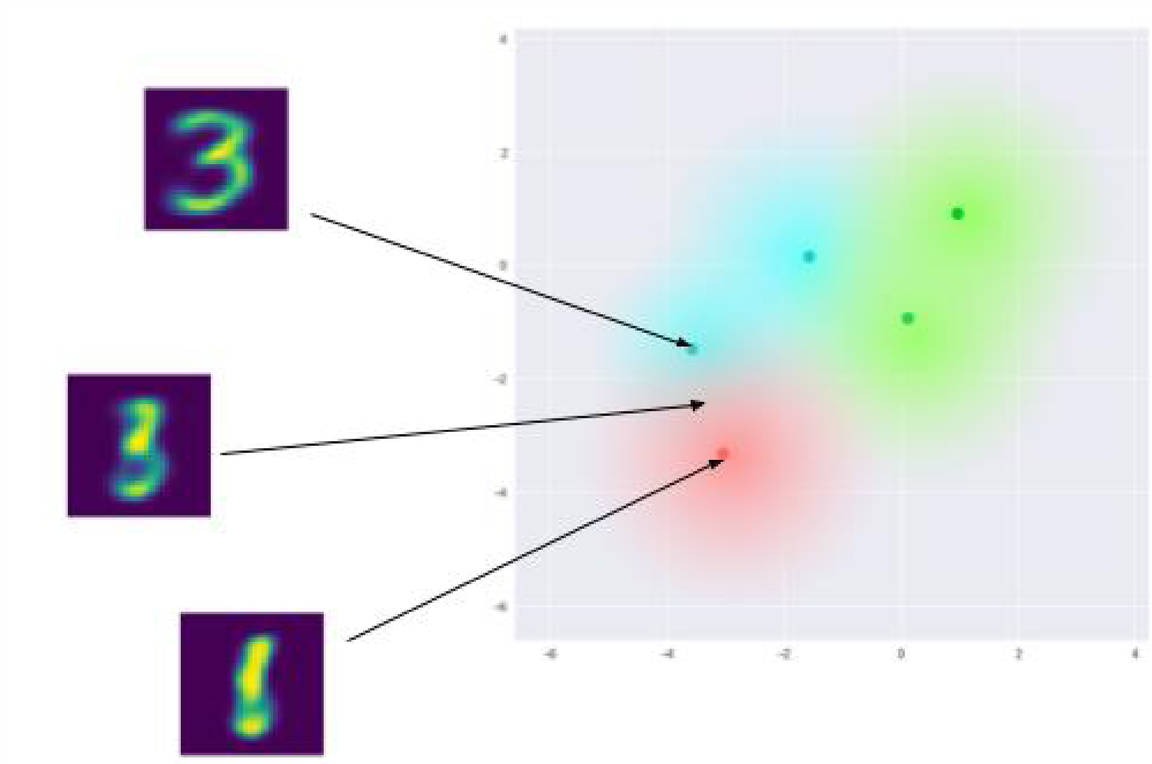
Расположение объектов в скрытом пространстве AE

Вариационные автокодировщики

В VAE каждый элемент скрытого пространства занимает не какую-то точку, а область (которая может пересекаться с областями других объектов).



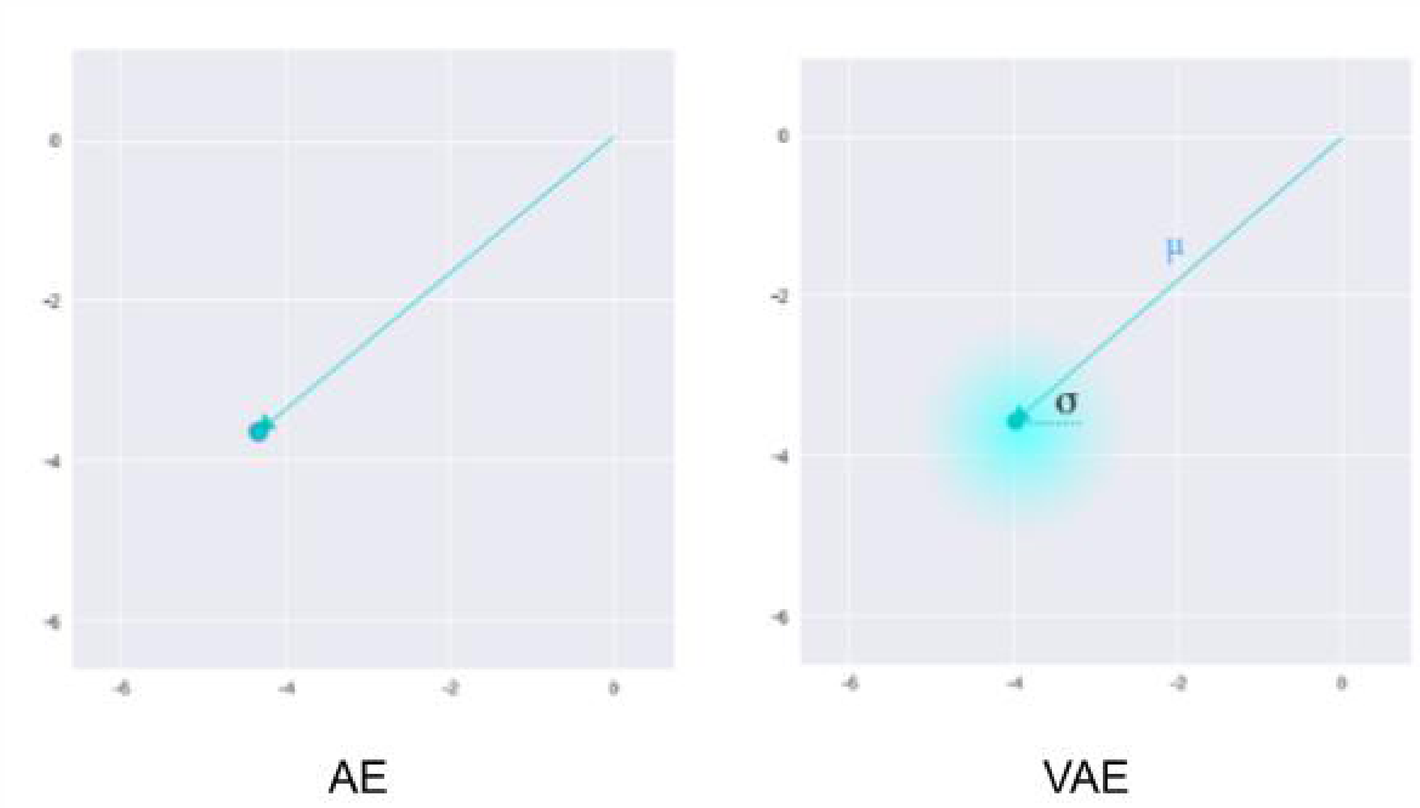
Благодаря «пересечениям» объектов внутри скрытого пространства мы можем получить какие-то новые объекты (на основе тех, которые пересекаются).



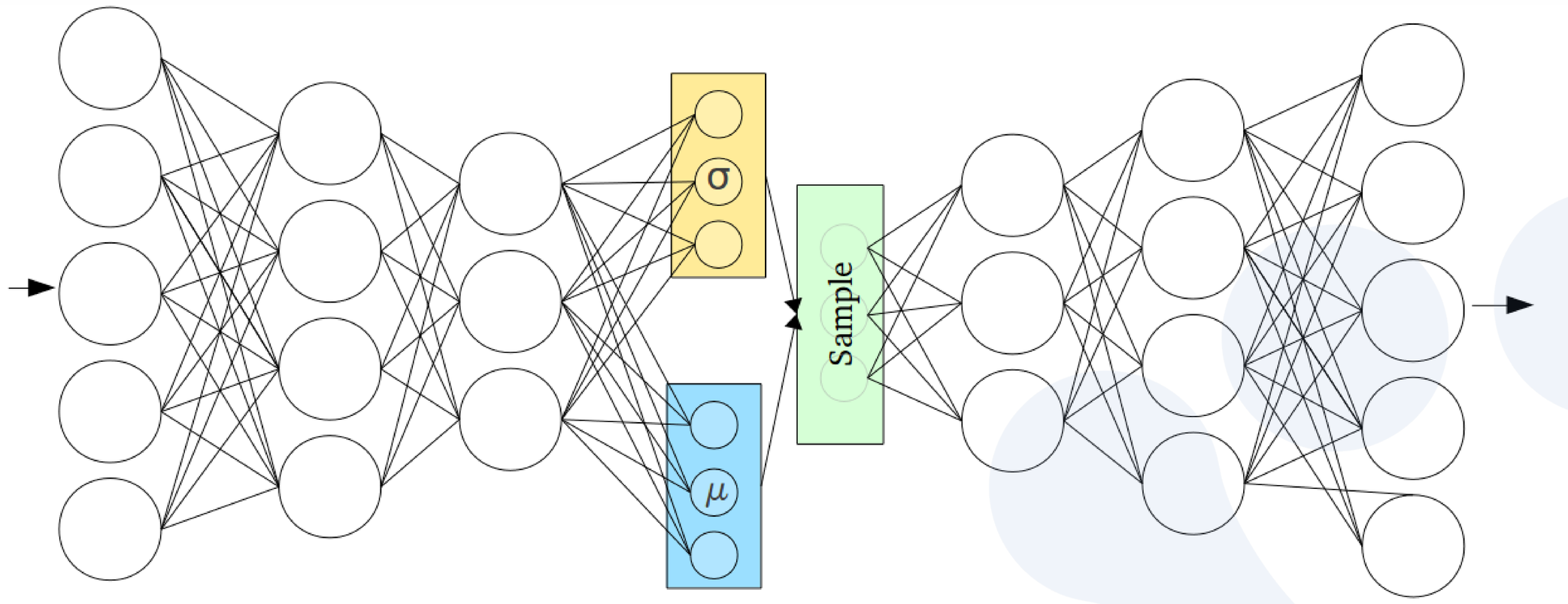
Например, если мы возьмем точку на пересечении объектов 1 и 3, то получим что-то среднее между ними.

Непрерывность скрытого пространства достигается следующим способом: энкодер выдает не один вектор размера n, а два вектора размера n – вектор средних значений μ и вектор стандартных отклонений σ.

Вариационные автокодировщики



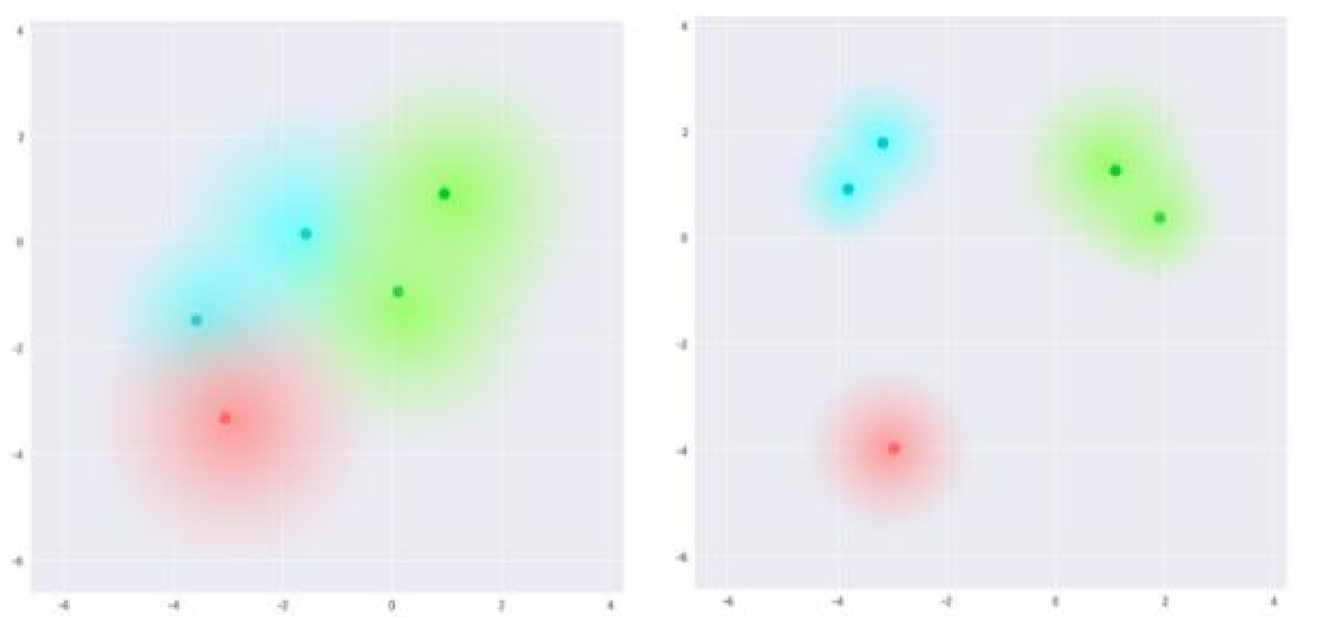
Среднее значение вектора определяет точку, вблизи которой будет вершина вектора, а стандартное отклонение определяет, насколько далеко может отстоять вершина от этого среднего. Таким образом, вершина вектора кодирования может лежать внутри n-мерного круга (см. рисунок выше). Входному объекту соответствует уже не одна точка в скрытом пространстве, а некоторая непрерывная область. Этот факт позволяет декодеру работать не с одним единственным вектором кодирования, соответствующим входным данным, а с их набором, благодаря чему в восстановление даже одного изображения вносится доля вариативности.



Теперь модель обладает вариативностью даже в пределах одного вектора кодирования, так как скрытое пространство локально непрерывно, т. е. непрерывно для каждого образца входных данных. В идеальном случае нам хотелось бы перекрытия этих локальных областей и для образцов входных данных, которые не похожи друг

Вариационные автокодировщики

на друга, чтобы производить интерполяцию. Однако так как нет ограничений на значения, принимаемые векторами μ и σ, энкодер может быть обучен генерировать сильно отличающиеся μ для разных образцов входных данных, тем самым удаляя их представления друг от друга в скрытом пространстве. Кроме того, энкодер будет минимизировать σ для того, чтобы векторы кодирования не сильно отличались для одного образца. Таким образом, декодер получает данные с малой степенью неопределенности, что позволяет ему эффективно восстанавливать данные из тренировочных сетов, но при этом мы можем не иметь непрерывного пространства.



Что нам нужно

Что мы можем непреднамеренно получить

Мы хотим, чтобы все области в скрытом пространстве были как можно ближе друг к другу, но при этом оставались различимыми как отдельные составляющие. В этом случае мы можем производить гладкую интерполяцию и создавать новые данные на выходе.

Для того чтобы достичь этого, в функцию потерь вводится так называемая **Kullback–Leibler расходимость**.

KL расходимость между двумя функциями распределения показывает, насколько сильно они отличаются друг от друга.

$$\sum_{i=1}^{n} \sigma_i^2 + \mu_i^2 - log(\sigma_i) - 1$$

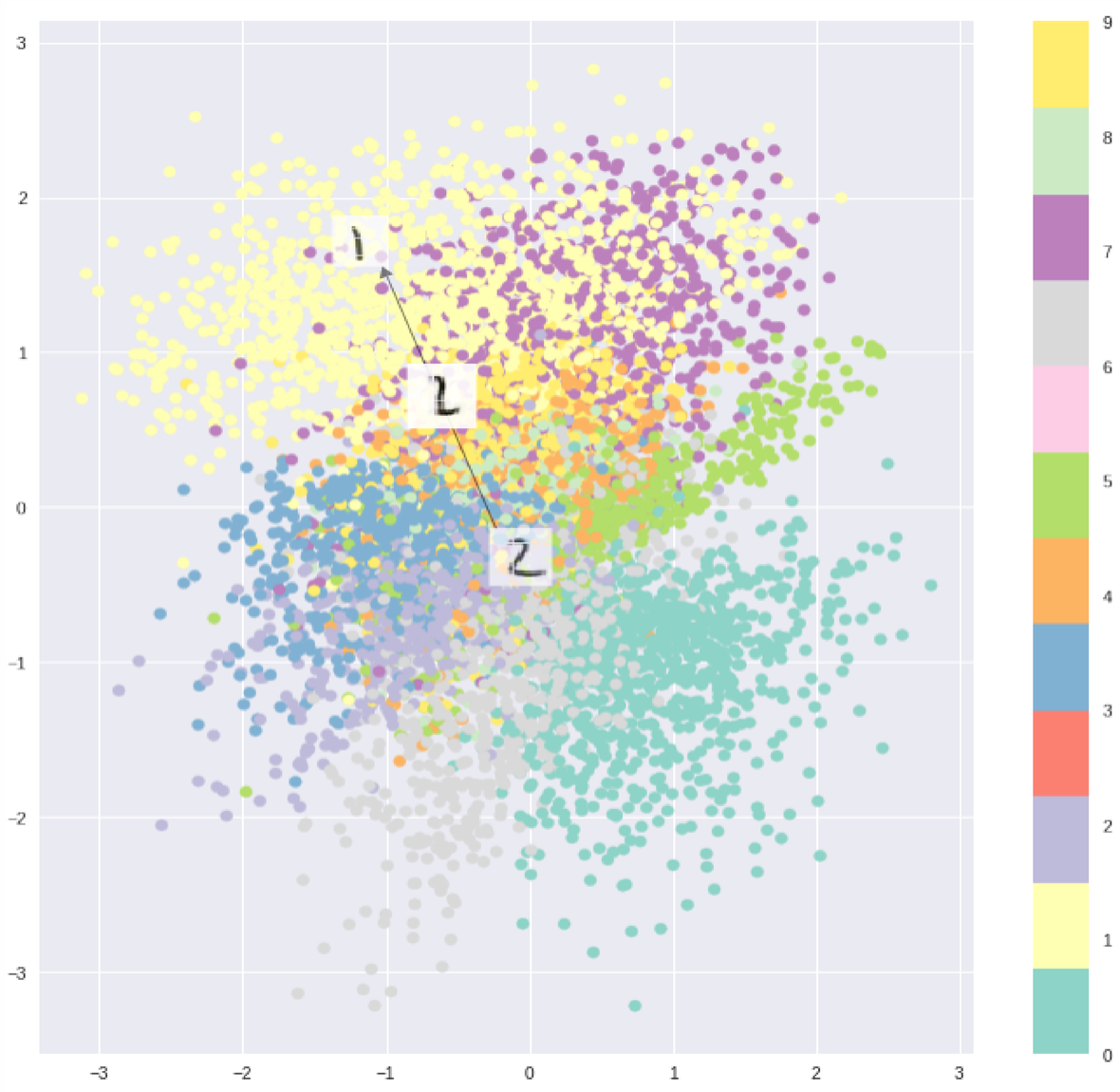
Учёт KL потерь заставляет энкодер помещать каждую отдельную область кодирования в окрестности некоторой точки в скрытом пространстве.

Вариационные автокодировщики

Ошибка Loss при обучении нейросети будет рассчитываться по следующей формуле:

$$0.5(tr(\sum_1^{-1}\sum_0) + (\mu_1 - \mu_0)^T \sum_1^{-1}(\mu_1 - \mu_0) - k + ln(\frac{det\sum_1}{det\sum_0})) + rmse$$

Оптимизируя и энкодер, и декодер, мы получаем скрытое пространство, которое отражает схожесть соседних векторов на глобальном уровне и имеет вид плотно расположенных областей возле начала координат скрытого пространства:



Достигнутый результат – это компромисс между кластерной природой потерь восстановления, необходимой декодеру, и нашим желанием иметь плотно расположенные векторы при использовании KL потерь.

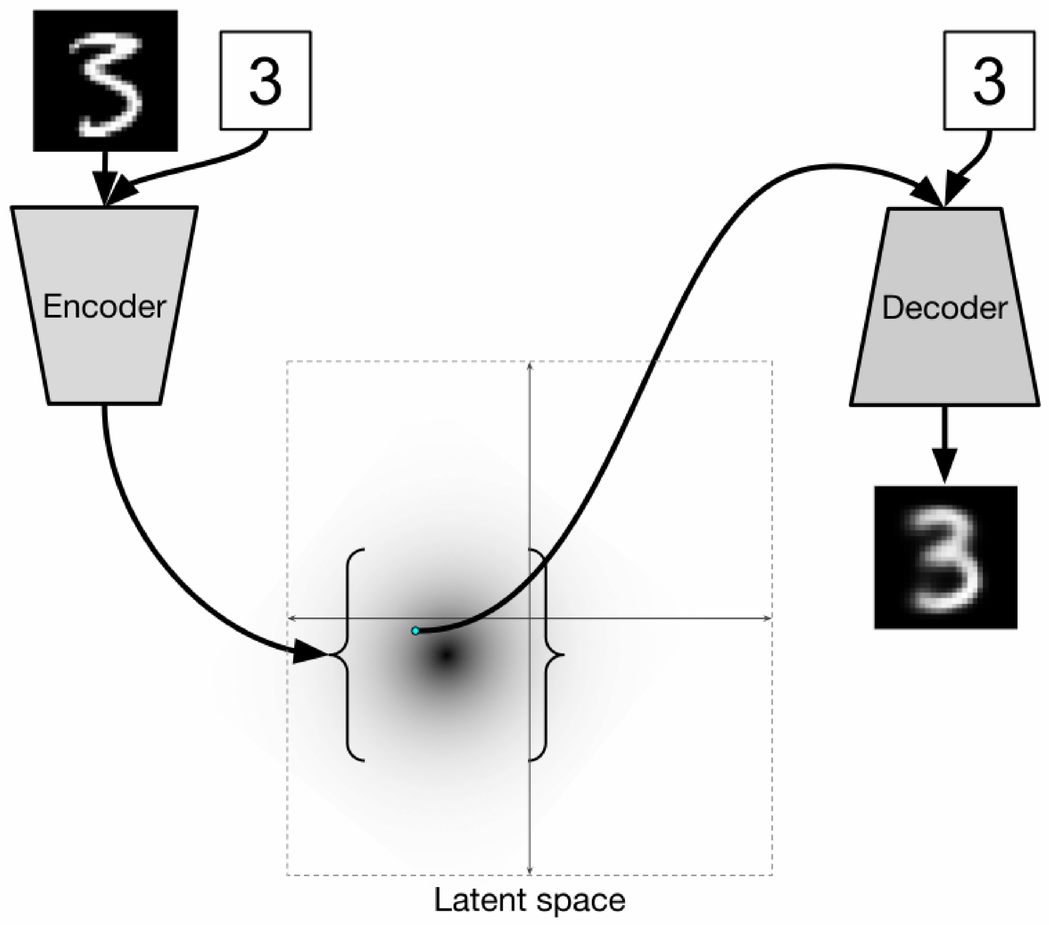
Минусы VAE:

- Нет гарантий, что в промежутках между областями, в которых были сконцентрированы варианты одного и того же объекта, находятся осмысленные изображения.
- Сложно генерировать картинку какого-то заданного объекта. Для этого необходимо смотреть, в какую область латентного пространства попадали изображения конкретного объекта.

Вариационные автокодировщики

Вариационные автокодировщики с условием (CVAE)

Если же теперь взять **VAE**, как в предыдущей части, и подавать на вход еще и лейблы, то получится **Conditional Variational Autoencoder (CVAE).**



Принцип построения CVAE следующий:

1. Обучаем CVAE на картинках с лейблами.
2. Кодируем стиль заданной картинки в латентное пространство.
3. Меняя лейблы, получаем из каждого вектора латентного пространства новые картинки.