

Методы векторного представления глубоких генеративных моделей

Мария Александровна Никитина

Московский физико-технический институт

Кафедра: Интеллектуальный анализ данных

Научный руководитель: кандидат ф.-м. наук О. Ю. Бахтеев

Научный консультант: А. Ю. Бишук

2024

Задача векторного описания генеративных моделей

Задача

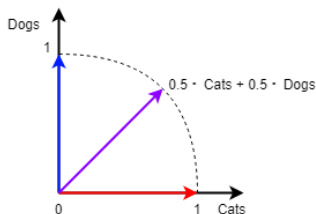
Задан набор генеративных моделей, описывающий разные выборки/генеральные совокупности данных. Требуется предложить метод векторного представления этих моделей, который будет представлять их статистические свойства.

Требования

1. Расстояние между векторными представлениями моделей для близких выборок должно быть невелико (при условии, что сами модели хорошо их описывают)
2. Модели, обученные на композиции/смеси выборок должны учитывать свойства всех выборок, входящих в смесь

Примеры возможных реализаций

1. Один из возможных вариантов: сумма векторных представлений моделей, полученных по датасетам D_1 , D_2 должна приблизительно соответствовать векторному представлению датасета $D_1 + D_2$;



2. Вместо использования евклидового расстояния на векторных представлениях, использовать иерархию;

3. Представить модель как граф и работать с пространством графов.

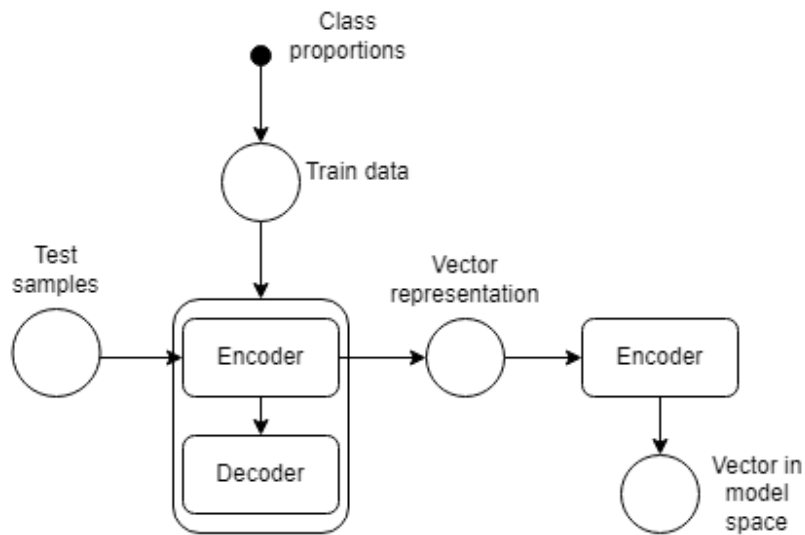
Описание эксперимента

Для базового эксперимента берётся 3 наиболее удалённых класса из датасета CIFAR. Для поиска таких классов использовалось евклидово расстояние на эмбедингах, полученных из выходного слоя ResNet.

Алгоритм

1. Случайное сэмплирование долей классов в датасете для n моделей;
2. Обучение n моделей на соответствующих датасетах;
3. Получение векторов из обученных моделей;
4. Обучение энкодера на полученных моделях. Предсказание:
 - 4.1 Вектора на части единичной сферы
 - 4.2 Расстояния между двумя моделями

Схема эксперимента



Возможные метрики

1. Contrastive N-pair loss, где $\text{Encoder}(m_i) = \mathbf{x}$, $\mathbf{d}_i = \mathbf{x}_i^+$, остальные элементы батча: \mathbf{x}_i^- :

$$\mathcal{L}_{N\text{-pair}}(f) = -\log \frac{\exp(\mathbf{x}^T \mathbf{x}_i^+)}{\exp(\mathbf{x}^T \mathbf{x}_i^+) + \sum_{i=1}^{N-1} \exp(\mathbf{x}^T \mathbf{x}_i^-)};$$

2. Угол между вектором модели $\text{Encoder}(m_i)$ и датасета \mathbf{d}_i :

$$\mathcal{L}_{\cos}(\text{Encoder}(m_i), \mathbf{d}_i) = \cos(\text{Encoder}(m_i) \cdot \mathbf{d}_i);$$

3. Triplet loss:

$$\mathcal{L}_{\text{Triplet}} = \sum_{\mathbf{x} \in \chi} \max(0, \|\mathbf{x} - \mathbf{x}^+\|_2^2 - \|\mathbf{x} - \mathbf{x}^-\|_2^2 + \epsilon;$$

4. MSE между $\text{Encoder}(m_i)$ и \mathbf{d}_i :

$$\mathcal{L}_{\text{MSE}}(\text{Encoder}(m_i), \mathbf{d}_i) = \|\text{Encoder}(m_i) - \mathbf{d}_i\|_2^2.$$

Результаты первого эксперимента

Способ представления модели:
среднее значение скрытого пространства автоэнкодера
на тестовой выборке.

Метрика энкодера: MSE

Проблемы: Переобучение
энкодера и неинформативность
взятия среднего

