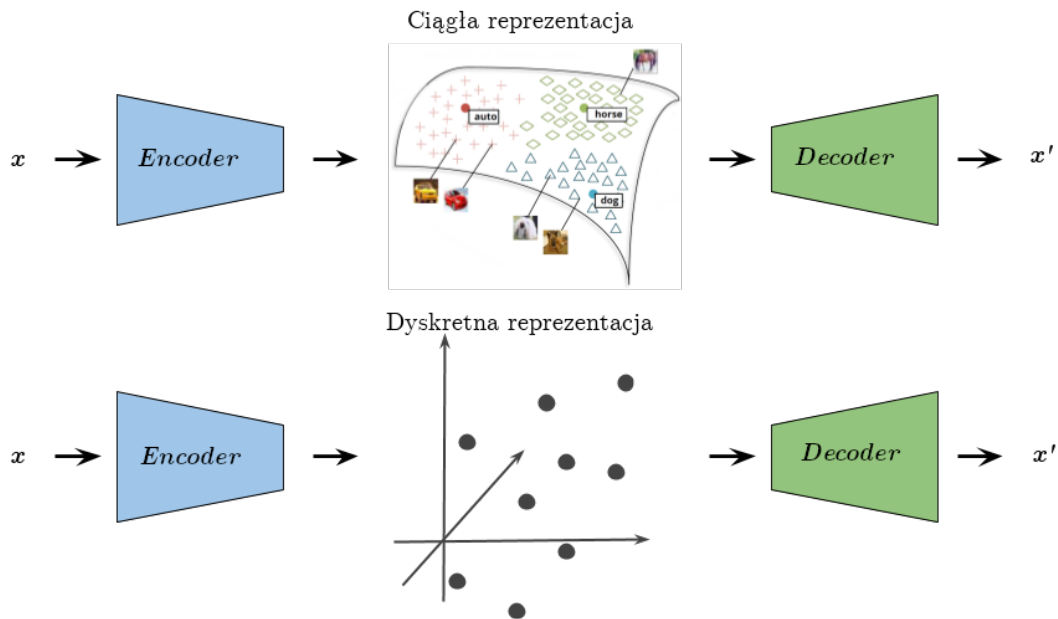


Rzadkie i dyskretne reprezentacje w ukrytych przestrzeniach

Uczenie się reprezentacji bez nadzoru jest bardzo ekscytującym kierunkiem w uczeniu maszynowym, szczególnie biorąc pod uwagę mnóstwo łatwo dostępnych danych bez etykiet. Istnieje wiele algorytmów uczenia maszynowego (np. wykrywanie obiektów, klasyfikacja, uczenie się ze wzmocnieniem, kompresja modelu lub generowanie nowych próbek), które mogłyby znacznie skorzystać na reprezentacji niskowymiarowej i wysoce wyrazisty cechach. Rzadkie i dyskretne reprezentacje zapewniają te właściwości.

Takie reprezentacje są bardziej interpretowalne niż ciągłe i gęste, które są zwykle tajemnicze, a zrozumienie ich zachowania jest niezwykle trudnym zadaniem. W sytuacjach, w których ludzie muszą podejmować decyzje w oparciu o duże ilości danych, interpretowalność ma zasadnicze znaczenie dla ułatwienia dalszych czynności podejmowanych przez człowieka. Obecnie coraz więcej wysiłku koncentruje się na tworzeniu modeli, które dostarczają zrozumiałych dla człowieka uzasadnień ich wyników. Przykład różnicy między ciągłą, gęstą reprezentacją, a dyskretną, rzadką reprezentacją przedstawiono na rysunku 1. Widzimy, że dyskretne i rzadkie reprezentacje dają nam znacznie lepsze możliwości interpretacji tego modelu.



Rysunek 1: Rysunek przedstawia dwa autoenkodery: na górze z ciągłymi zmiennymi ukrytymi, na dole z dyskretnymi. Różnica między nimi polega na tym, że w górnym modelu każde wejście x ma swoją własną reprezentację w utajonej przestrzeni, a w drugim modelu reprezentacje w utajonej przestrzeni x można zdefiniować przez kombinację wszystkich dyskretnych zmiennych (ich liczba jest znacznie mniejsza niż w przypadku wyższego modelu).

Dyskretne i rzadkie reprezentacje są również znacznie bardziej intuicyjne i solidne niż ciągłe i gęste. Potwierdza to wiele biologicznych inspiracji, a także eksperymentów na sztucznych sieciach neuronowych. Ponieważ wiele architektur nadal opiera się na ciągłych i gęstych reprezentacjach, istnieje potrzeba zgłębienia tego tematu i opracowania nowych podejść.

Projekt uważamy za istotny ze względu na jego potencjalny pozytywny wpływ na następujące problemy:

- *reprezentację danych* – duża liczba zbiorów danych zawiera z natury dyskretne czynniki generujące, które są trudne do wykrycia przez sieci neuronowe z ciągłymi zmiennymi;
- *interpretowalność* – zmienne dyskretne są łatwiejsze do zinterpretowania, a rzadka reprezentacja obejmuje sam proces generowania danych i jest inspirowana biologicznie;
- *wydajność obliczeniowa* – przetwarzanie rzadkich i dyskretnych danych wymaga mniej pracy i energii.

Tworzenie użytecznych, rzadkich i dyskretnych reprezentacji w ukrytych przestrzeniach jest bardzo trudnym tematem badawczym. W naszym projekcie chcemy rozwiązać ten problem. Chcemy zaproponować nowe metody tworzenia dyskretnych i rzadkich reprezentacji. Użyć je do rozwiązania problemów, w których ich zastosowanie będzie nowatorskie i może znacznie poprawić wyniki. Planujemy przeanalizować, jak rzadka i dyskretna reprezentacja może pomóc w tworzeniu samowyjaśniających się modeli. Zastosujemy nasze metody do różnych bieżących wyzwań uczenia maszynowego, takich jak kompresja dużych modeli (np. model BERT). Wierzymy, że nasze propozycje nie tylko przewyższą dokładność już istniejących modeli, ale także poprawią ich interpretowalność.