

Serwis reprezentatywności

W ramach zadania należy zaimplementować serwis REST'owy w języku Python, który pozwoli na wyuczenie i następnie odpytywanie modelu o tak zwaną reprezentatywność obiektu (szczegóły w dalszej części polecenia).

Serwis ten powinien udostępniać następujące endpoint'y:

1. Zlecenie wytrenowania nowego modelu na przesłanych danych.
2. Sprawdzenie statusu ostatnio zleconego treningu.
3. Predykcja ostatnio wytrenowanego modelu na przesłanych obiektach.

Miarę reprezentatywności obiektu rozumiemy jako liczbę, która opisuje to jak dobrze dany obiekt reprezentuje inne obiekty w danym zbiorze. Im większa jest reprezentatywność danego obiektu tym więcej w opisywanej przestrzeni danych jest obiektów podobnych.

Istnieje wiele sposobów na wyliczanie reprezentatywności. W tym zadaniu należy zaimplementować tylko jeden z nich, który przebiega następująco:

1. Dzielimy zbiór danych na L mniej więcej równych części w sposób losowy.
2. Każdą część zbioru będziemy przetwarzali współbieżnie:
 - a. Liczymy odległość każdego obiektu z danego podzbioru od K najbliższych obiektów w danym fragmencie zbioru.
 - b. Obliczamy średnią z tych k wielkości.
 - c. Reprezentatywność obiektu definiujemy jako:
$$\frac{1}{1 + \text{średnia_odległość_k_sąsiadów}}$$
 - d. Gdy mamy mapowanie obiekt \rightarrow reprezentatywność, uczymy model, który będzie w stanie przewidywać reprezentatywność na nowych danych.
3. Zbieramy wszystkie L wyliczonych modeli, nasz finalny model będzie ich ensemble'm zwracającym średnią z ich predykcji.

Ad 1. REST powinien pozwalać na wyuczenie nowego modelu według opisu powyżej, podczas uczenia serwis dalej powinien być responsywny, w szczególności zwracać informację czy zadany trening już się zakończył. Jako dane wejściowe endpoint powinien przyjmować listę reprezentacji obiektów. Zakładamy, że każdy obiekt będzie reprezentowany przez listę liczb rzeczywistych o tej samej długości.

Ad 2. REST powinien zwracać jeden z 3 następujących statusów:

- Błąd podczas uczenia wraz z krótkim opisem błędu, timestampem początku uczenia i wystąpienia błędu
- Uczenie nadal trwa wraz z czasem rozpoczęcia
- Uczenie zakończone wraz z czasem początku i końca treningu

Ad 3. Dla zadanej listy obiektów, REST powinien zwrócić listę reprezentatywności. Gdzie kolejne reprezentatywności są wartościami zwracanymi przez ostatnio wytrenowany model dla odpowiadających im obiektów.