KARTA PRACY UCZESTNIKA MODUŁ 3 ROZDZIAŁ 3

Podobieństwo wektorów

Od prostych wektorów do porównania pytania z trzema fragmentami

Czas orientacyjny	Pracujesz z	Rezultat
25 min	NumPy, funkcją cosine_similarity() i wynikiem R2	podobienstwa.csv i własna interpretacja

Sprawdzisz, co opisuje podobieństwo kosinusowe i jak różni się od oceny przydatności tekstu. Wykorzystasz gotowe embeddingi fragmentów z rozdziału 2.

Cel i rezultat rozdziału

Obliczysz i zinterpretujesz podobieństwo pytania do fragmentu, uwzględniając błędy danych.

Po zakończeniu masz

- funkcję cosine_similarity() sprawdzoną na prostych wektorach;
- tabelę trzech porównań z rzeczywistymi chunk_id;
- wniosek wyjaśniający różnicę między score a poprawnością.

Zanim zaczniesz

Najpierw wykonaj notebook 00. Otwórz wskazany starter, wybierz MODE i uruchom Przygotowanie. Pola UZUPEŁNIJ wypełnij samodzielnie. SOURCE wybiera własne R2 albo checkpoint; nie ma automatycznej zamiany wejścia.

```
notebooki/M12_M3_R3_kosinus_start.ipynb
```

Materiał	Do czego służy
chunks_with_embeddings.jsonl z R2	Trzy wybrane rekordy ze zgodnymi wektorami.
embedding_manifest.json z tej samej próby	Model, wymiar i pochodzenie wektorów.
NumPy i lokalny model	Obliczenia oraz embedding nowego pytania.

Gdzie wrócić w prezentacji

W R3 odszukaj „Funkcja podobieństwa w NumPy” i „Porównanie pytania z pojedynczym fragmentem”. Definicja create_embedding() jest w R2 na slajdzie o przygotowaniu Ollama.

Praca krok po kroku

Krok 1 Sprawdź geometrię — Przewidź podobieństwo par [1, 0] z [1, 0], [0, 1] i [-1, 0].

- Zapisz przewidywania przed obliczeniem.
- To wektory kontrolne, nie embeddingi tekstu z BGE-M3.

Krok 2 Sprawdź funkcję — Przeczytaj cosine_similarity(vector_a, vector_b) oraz normalize_vector() w kod/m3_tools.py.

- Sprawdź kontrolę wymiarów, skończoności i norm; uruchom próby błędnych danych.
- Wariant własny: napisz osobną funkcję i porównaj wyniki z implementacją referencyjną.

Krok 3 Wykonaj kontrole — Oblicz wyniki dla prostych par i porównaj je z przewidywaniami.

- Sprawdź również wektor zerowy oraz różne wymiary.
- Wyjaśnij, co zmienia normalizacja długości wektora.

Krok 4 Wybierz trzy fragmenty — Dla jednego pytania wybierz fragment trafny, częściowo związany i niezwiązany.

- Uzupełnij wybor i uzasadnienia w notebooku przed obliczeniem score.
- Użyj rzeczywistych identyfikatorów z wybranego pliku R2.

Krok 5 Porównaj embeddingi — Utwórz embedding pytania i porównaj go z zapisanymi wektorami fragmentów.

- Użyj model_name=EMBEDDING_MODEL; nie przeliczaj niezmienionych fragmentów.
- Zapisz question, chunk_id, score, ocenę przydatności i uzasadnienie.

Krok 6 Zapisz interpretację — Uporządkuj trzy wyniki i porównaj ranking z oceną tekstu.

- W podobienstwa.csv zapisz pełny wynik liczbowy; zaokrąglaj tylko widok.
- We wniosku wyjaśnij, dlaczego wynik 0,82 nie oznacza 82% pewności odpowiedzi.

Zapis próby: komórka Przygotowanie automatycznie tworzy nowy folder wyniki/R3/<proba_id>/. Zachowaj wyświetloną ścieżkę.

Punkty kontrolne pracy

Kiedy	Co sprawdzasz	Co zrobić przy problemie
Proste wektory	Wyniki są zgodne z geometrią kierunków.	Sprawdź iloczyn skalarny i normy.
Błędy danych	Zero i różne wymiary dają jawny błąd.	Uzupełnij walidację funkcji.
Trzy fragmenty	Tabela łączy score z tekstem i chunk_id.	Porównaj wynik z rzeczywistą treścią.

Nazewnictwo i zapis wyników

Notebook zapisuje podobienstwa.csv i wnioski.md w wyniki/R3/<proba_id>/. W notatniku zachowaj definicję funkcji i wyniki kontroli.

Identyfikator <proba_id> nadaje notebook. Pliki są chronione przed nadpisaniem; nowy pomiar rozpocznij od Przygotowania. Zapisz notebook przez Ctrl+S i wskaż próbę do oceny.

Lista kontroli przed zakończeniem

- □ Zapisałem przewidywania przed pomiarem.
- □ Sprawdziłem proste wektory i błędy danych.
- □ Pytanie i fragmenty mają zgodny model.
- □ Wybrałem trzy rzeczywiste rekordy.
- □ Oddzieliłem score od oceny przydatności.
- □ Zapisałem tabelę oraz wniosek.

Gdy pojawi się problem

Objaw	Co sprawdzić
Niezgodny wymiar	Sprawdź manifest R2 i model embeddingu pytania.
Nieoczekiwana kolejność	Przeczytaj fragmenty i opisz różnicę; nie poprawiaj score ręcznie.
Brak danych po restarcie	Wczytaj wybraną próbę R2 i odtwórz funkcje.

Wynik kosinusowy opisuje relację wektorów. Przydatność fragmentu ustalasz na podstawie tekstu i pytania. Zapisz także uzupełniony notebook.