Dense vs Sparse

Sparse representation

- one-hot encoding is the example of sparse representation;
- a vector where most values are 0, and only one value is 1;
- example for a 10 000-word vocabulary: [0,0,1,0,0,...,0] (size 10,000);
- problems:
    - very high dimensionality;
    - no semantic relationships (all words are equally distant);
    - computationally inefficient.

Dense representations

- embeddings create dense vectors;
- compact, lower-dimensional vectors (e.g., 300 dimensions instead of 10 000);
- example [0.24, -0.87, 0.52, ..., 0.19] (size 300)
- advantages:
    - captures semantic relationships (similar words have similar vectors);
    - computationally efficient;
    - better generalization.

The typical workflow

Raw Text → One-Hot Encoding → Embedding Layer → Dense Vectors → Neural Network

Semantic Meaning

- similar words have similar vectors:
    - dog ≈ [0.1, 0.3, 0.4, ...];
    - puppy ≈ [0.2, 0.4, 0.3, ...] (close to dog);
    - car ≈ [0.8, 0.1, -0.2, ...] (far from dog).