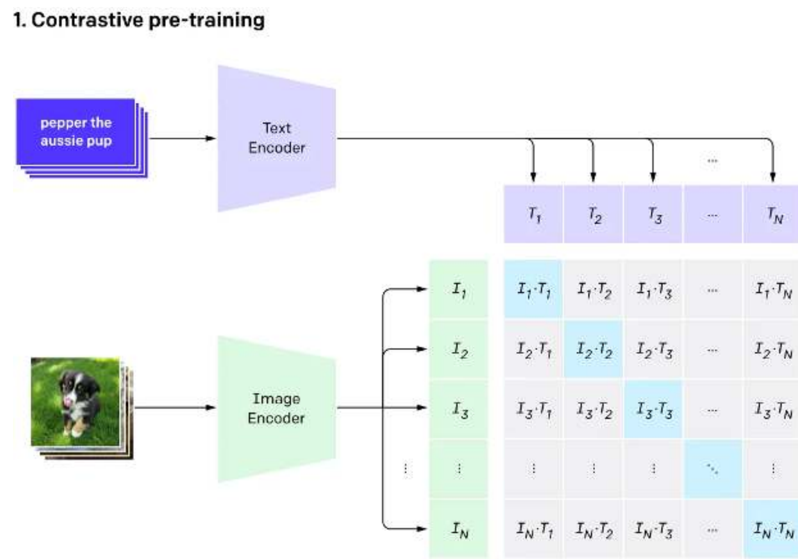
Semi Supervised Image Labelling with Vector Embeddings

Washington University in St. Louis

SCHOOL OF MEDICINE

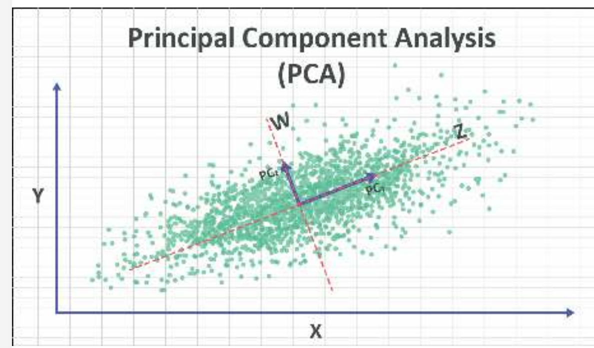
Intro to Image Embeddings

- Multimodal Learning: CLIP (Contrastive Language–Image Pretraining) is an AI model that learns visual concepts from natural language descriptions, enabling it to understand and generate embeddings for both images and text.
- Unified Embedding Space: Maps images and text into a shared space, positioning semantically similar items close together.
- Cosine Similarity: Uses cosine similarity to compare CLIP embeddings, identifying the most semantically similar items efficiently.
- Vector DB Integration: Stores CLIP embeddings in VectorDBs for scalable, high-performance similarity searches.
- Versatile Applications: Enables zero-shot classification, content-based filtering, and multimodal AI systems by linking visual and textual information.



Intro to Feature Decomposition with PCA

- Statistical technique for dimensionality reduction of large datasets
- Identifies principal components - directions of maximum variance in the data
- Transforms data into a new coordinate system based on these components
- First principal component captures the most variance, second captures second most, etc.
- Allows data compression by keeping only top N components
- Useful for visualization, noise reduction, and feature extraction in machine learning



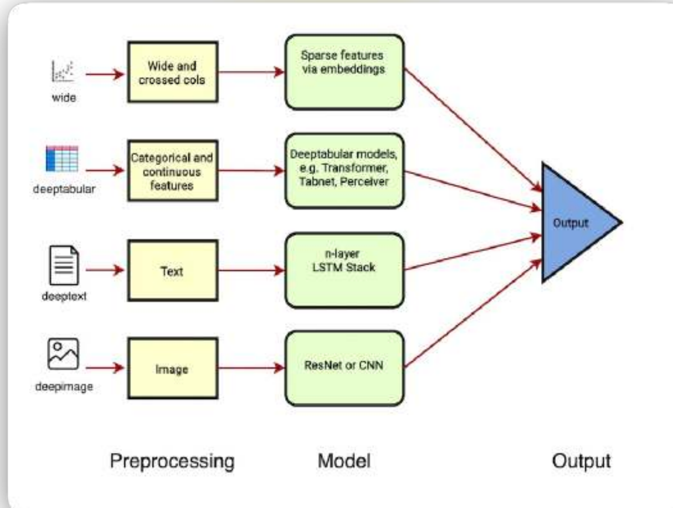
KMeans for Semi Supervised Labelling

- Initial Clustering: Apply KMeans to the entire dataset, including both labeled and unlabeled data, using features like CLIP embeddings.
- Label Propagation: Assign the majority label from labeled points to all points within each cluster, effectively "propagating" labels to unlabeled data.
- Confidence Thresholding: Set a confidence threshold based on cluster purity or distance from cluster centroid. Only propagate labels to unlabeled points exceeding this threshold.
- Iterative Refinement: Use newly labeled data to retrain a supervised model, then recluster and repeat the process, gradually improving classification accuracy.
- Active Learning Integration: Identify uncertain points (e.g., near cluster boundaries) for human labeling, focusing manual effort where it's most impactful.
- Ensemble Approach: Combine KMeans with other semi-supervised methods (e.g., label spreading, self-training) for more robust results, especially in complex datasets like medical imaging.

Let's CLIP it in the Bud

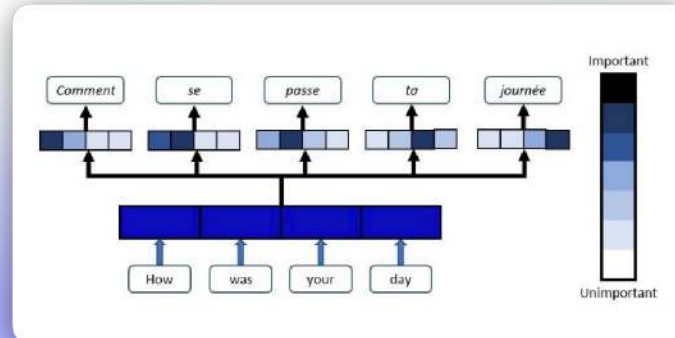
Introduction to Multimodal Deep Learning

- Multimodal learning combines various data types, such as images, text, and genomic data, for richer insights.
- Offers improved diagnostic accuracy and prediction robustness over unimodal methods by leveraging the complementary nature of different data types.
- Applications range from enhanced diagnostics to precise treatment planning and continuous patient monitoring.



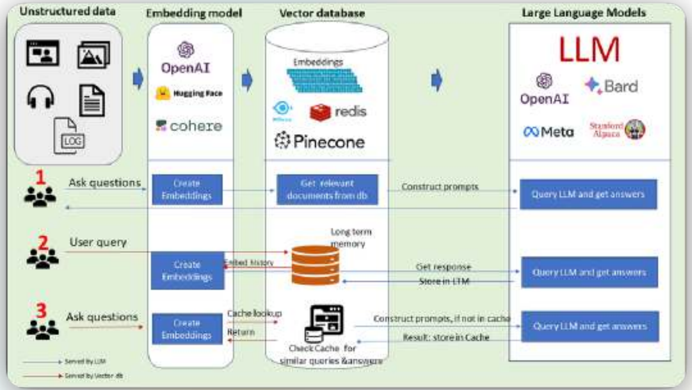
Transformers in Healthcare

- Transformer models excel in healthcare by analyzing multimodal data without sequential processing constraints.
- They efficiently handle varied healthcare data types, improving upon RNNs and CNNs.
- With self-attention mechanisms, transformers capture complex data dependencies, enhancing multimodal analysis.
- Their interpretability benefits healthcare decision-making, offering transparency in patient care analysis.
- Transformers' versatility makes them a promising tool for integrating diverse healthcare data, aiming to improve patient outcomes.



Vector Databases

- Efficient Similarity Search: VectorDBs enable efficient similarity searches for CLIP vector embeddings by indexing high-dimensional vectors, allowing quick retrieval of semantically similar items.
- Scalability: VectorDBs are designed to handle large-scale datasets, making them suitable for managing and querying extensive collections of CLIP embeddings generated from diverse and voluminous data sources.
- Real-time Applications: Content-based image retrieval, by quickly finding and serving relevant images or text based on the input CLIP embeddings.
- Integration with AI Pipelines: VectorDBs seamlessly integrate with AI pipelines, enhancing the performance of tasks like recommendation systems, visual search engines, and multimodal AI applications that rely on CLIP embeddings.



Summary

- Feature Representation: Use vector embeddings (e.g., CLIP) to convert raw data (like images) into high-dimensional feature vectors, capturing semantic information.
- Dimensionality Reduction: Apply PCA to reduce the dimensionality of the embeddings, retaining most important features while reducing noise and computational complexity.
- Clustering and Label Propagation: Perform KMeans clustering on the PCA-reduced embeddings. Propagate labels from labeled data points to unlabeled points within the same cluster, based on cluster purity or distance metrics.
- Iterative Refinement: Use the newly labeled data to train a classifier, then use this classifier to refine cluster assignments and label propagation. Repeat this process to improve labeling accuracy.
- Uncertainty Sampling: Identify data points near cluster boundaries or with low confidence scores for manual labeling, effectively combining active learning with the semi-supervised approach.
- CLIP and Vector Databases: Maps images and text into a shared space for classification and filtering, supported by scalable VectorDBs for similarity searches.