Can All Negative Pairs Be Orthogonal?

Due: Wednesday, November 18 at 11:59pm

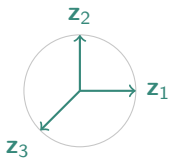
Instructions: Please submit your *handwritten* solutions as a single PDF on Gradescope.

For autoencoders, linear maps can limit reconstruction. Let's ask whether embedding dimension also limits contrastive learning. Let's use a simpler loss than cross-entropy: squared error with a target inner product of one for positive pairs and zero for negative pairs.

Suppose each of n positive pairs gets identical unit-length embeddings. Put one embedding per pair in the rows $\mathbf{z}_i^\top$ of $\mathbf{Z} \in \mathbb{R}^{n \times k}$, where $n > k \geq 1$. Treat examples from different pairs as negatives. The two batches then have $\mathbf{Z}' = \mathbf{Z}$, so the similarity matrix is $\mathbf{Z}\mathbf{Z}^\top$. Writing $\mathbf{I}_n$ for the $n \times n$ identity matrix, our loss is

$$\mathcal{L} = \|\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_n\|_F^2.$$

The encoder may be linear or nonlinear. We'll ask what the embedding dimension alone forces us to lose. Here are three unit vectors in the plane and their similarity matrix.



$$\mathbf{Z}\mathbf{Z}^\top = \begin{bmatrix} 1 & 0 & -1/\sqrt{2} \\ 0 & 1 & -1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} & 1 \end{bmatrix}$$

1. Each embedding has unit norm, so every positive pair already has the desired similarity. Show that the loss comes entirely from the negative pairs:

$$\mathcal{L} = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} (\mathbf{z}_i^\top \mathbf{z}_j)^2.$$

2. The matrix $\mathbf{Z}^\top\mathbf{Z}$ has only k rows and columns. Show that we can compute the same loss from this smaller matrix:

$$\mathcal{L} = \|\mathbf{Z}^\top\mathbf{Z}\|_F^2 - n.$$

3. Let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of $\mathbf{Z}^\top\mathbf{Z}$. Use their sum to prove

$$\mathcal{L} \geq \frac{n(n-k)}{k}.$$

Hint: For real numbers $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$, the Cauchy–Schwarz inequality states

$$\left(\sum_{j=1}^{k} a_j b_j \right)^2 \leq \left(\sum_{j=1}^{k} a_j^2 \right) \left(\sum_{j=1}^{k} b_j^2 \right).$$

Takeaway: *Even when every positive pair matches perfectly, limited embedding dimension prevents all negative pairs from being perpendicular. A more expressive encoder cannot remove this geometric limitation of the squared-error objective.*