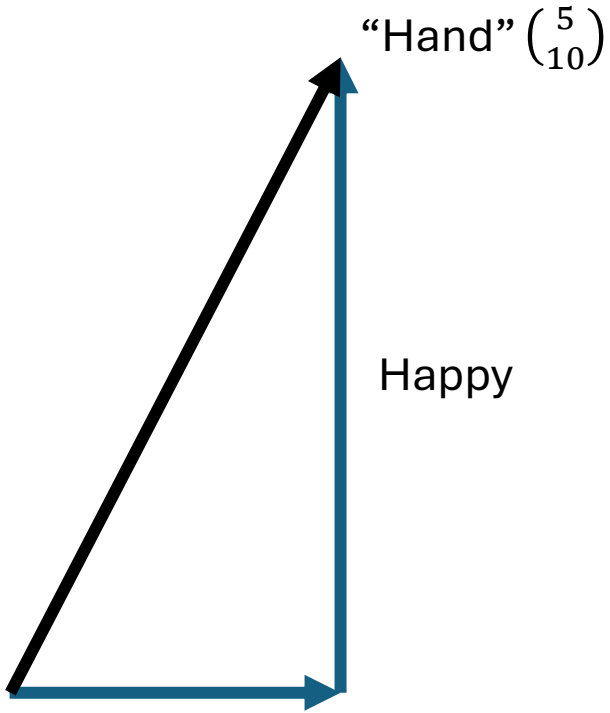

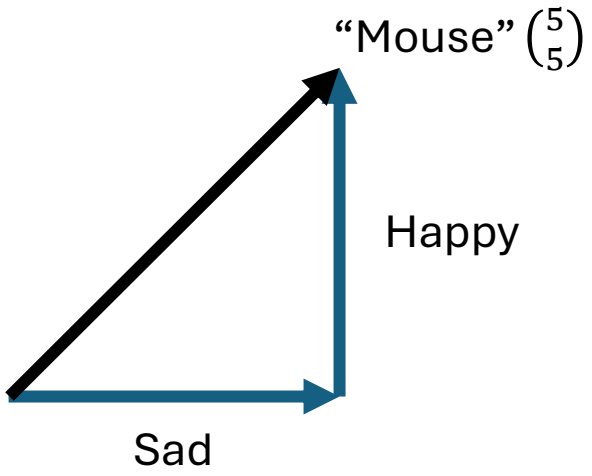

Intention:
Higher Attention Score = Higher Similarity

“Mouse” ------- “Hand”

$$\begin{pmatrix}5\\10\end{pmatrix}\cdot\begin{pmatrix}5\\5\end{pmatrix}=(5\times5)+(5\times10)=75$$

Relationship Between “Mouse” and “Hand”
Attention Score = 75

BUT
“Mouse” ---------- “Mouse”

$$\begin{pmatrix}5\\5\end{pmatrix}\cdot\begin{pmatrix}5\\5\end{pmatrix}=(5\times5)+(5\times5)=50$$

Relationship Between “Mouse” and “Hand”
Attention Score = 50

This is like raw data, we need to normalize to handle outliers and anomalies to prevent issues like this.