**3. Contestualizzazione con il Transformer (self-attention)**

**Esempio semplificato del procedimento:**

**1. Calcoliamo Q, V, K per i token “Bitcoin” e “raggiunge”**

Supponimao che gli embedding iniziali abbiano dimensione d = 3 e che scegliamo = 2.

Quindi le matrici di pesi saranno (pesi ipotizzati):

E il token “Bitcoin” ha embedding iniziale (dopo lookup e posizioni):

Allora:

Mentre il token “raggiunge” ha embedding iniziale

Allora:

**2. Calcoliamo la compatibilità**

Prendiamo la Query di “raggiunge”:

La confrontiamo con le key di “Bitcoin” e “raggiunge”:

* compatibilità con “Bitcoin”:
* compatibilità con “raggiunge”:

**3. Normalizziamo con softmax**

:

* Bitcoin: 1.2741 / 1.41 = 0.903
* Raggiunge: 2.1234 / 1.41 = 1.507

Applichiamo softmax:

**4. Costruiamo la nuova rappresentazione**

La nuova rappresentazione di “raggiunge” sarà una combinazione pesata di:

Il nuovo vettore di **“raggiunge”** non è più solo un verbo generico, ma porta dentro parte del significato di “Bitcoin” (peso 0.354) e parte di sé stesso (peso 0.646). Se avessimo incluso anche “record”, molto probabilmente avrebbe preso **ancora più peso**, perché “raggiungere un record” è un legame semantico forte.

In parallelo, anche “Bitcoin” e gli altri token aggiornano la propria rappresentazione. Alla fine otteniamo una sequenza [t1​,…,tm​] di vettori contestualizzati, in cui ciascun token incorpora il contesto degli altri. Questo processo si ripete per più layer e ad ogni passo fa emergere relazioni semantiche più ampie.

**C) Pooling di frase (mean pooling)**

**Esempio semplificato:**

Applicando il mean pooling ai vettori dei due token ottenuti precedentemente tb e tr abbiamo:

Questo vettore s è l’embedding della frase *“Bitcoin raggiunge”* nello spazio R2 ( in reltà sarebbe R384 ma lo abbiamo ridotto per semplicità).

Questo è l’**embedding di frase**, cioè il vettore che rappresenta l’intero paragrafo o titolo e che può essere confrontato con altri tramite la similarità coseno.