

# Homework 8

Anshul Sawant

March 22, 2025

## Contents

<b>1</b>	<b>Q 1.1</b>	<b>1</b>
1.1	B . . . . .	1
<b>2</b>	<b>Q 1.4 Cross-encoding</b>	<b>1</b>
2.1	A . . . . .	1
2.2	B . . . . .	2
2.3	C: Strategy to use both . . . . .	2

## 1 Q 1.1

### 1.1 B

It is sounder to use last-token pooling with decoder-only models because decoders process tokens left to right and a decoder would not have processed all of the input before the last token.

## 2 Q 1.4 Cross-encoding

### 2.1 A

```
def cross_encoder(query, passage):  
    reprs = encoder(query + '[SEP]' + passage)  
    similarity = Linear(reprs)  
    return similarity
```

## 2.2 B

**Advantage:** Higher accuracy because jointly considering query and passages allows for more nuanced semantic relationships. **Disadvantage:** Each query-passage pair requires a forward pass of the encoder making it more expensive. A bi-encoder will require  $|Q| + |P|$  forward passes whereas, a cross-encoder will require  $|Q| \times |P|$  forward passes.

## 2.3 C: Strategy to use both

A candidate set can be retrieved using bi-encoder and then this smaller set can be re-ranked using cross-encoder.