

# Lecture Notes

## Attention

### Query vs. Keys vs. Values

Query = what do you want to know?

Keys = labels for the values (to be matched with the queries)

Values = values.

How it works: Query \* Keys (dot product) tells how well each label works for your query. After a softmax layer, take the sum of this with the values.

### Self-Attention

Queries, keys, and values are all from the same source, but just different projections. I.e., if they started out as  $R^N$ , you might get  $R^k$ ,  $R^k$ , and  $R^v$  outputs.