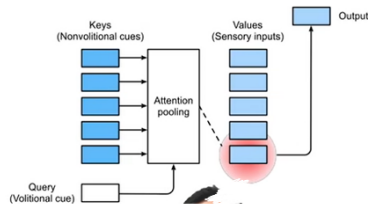


# 注意力机制

- 动物需要在复杂环境下有效关注值得关注的点
- 心理学框架：人类根据随意（刻意，主观）线索和不随意线索选择注意点

## 注意力机制

- 卷积，全连接，池化都只考虑不随意线索
- 注意力机制则显示的考虑随意线索
  - 随意线索称为查询（query）
  - 每个输入（环境）是一个值（value）和不随意线索（key）的对
  - 通过注意力池化层来有偏向性选择某些输入



原有的 key-value 对在 query 的影响下  
形成了新的 value  
不同的 query 会产生不同的 value

## 非参注意力池化层

- 给定数据  $(x_i, y_i), i=1, \dots, n$
- 平均池化是最简单的方案： $f(x) = \frac{1}{n} \sum y_i$  相当于对 query  $x$ ，直接返回所有 value 均值
- 更好的方案是60年代提出的Nadaraya-Watson核回归

$$f(x) = \sum_{i=1}^n \frac{k(x-x_i)}{\sum_{j=1}^n k(x-x_j)} y_i \rightarrow \text{value}$$

query
key
K是核函数，用于衡量  $x$  和  $x_i$  距离
选择与  $x$  相近的 key 给予它们 value 更大权重

使用高斯核时  $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$

$$f(x) = \sum_{i=1}^n \frac{\exp(-\frac{1}{2}(x-x_i)^2)}{\sum_{j=1}^n \exp(-\frac{1}{2}(x-x_j)^2)} y_i = \sum_{i=1}^n \text{softmax}(-\frac{1}{2}(x-x_i)^2) y_i$$

## 参数化注意力机制

- 在之前基础上引入可学习的  $w$

$$f(x) = \sum_{i=1}^n \text{softmax}(-\frac{1}{2}(x-x_i)w)^2 y_i$$