

- ML常见函数的作用

1. 处理过载信息，可以用注意力机制和优化网络结构
2. 注意力机制可以分为自上而下的聚焦式注意力，自下而上的显著式注意力

最大型卷积和门控神经网络可以deem as显著式注意力

consider 自上而下的聚焦式注意力，

examine n是否在X中，

$$\alpha_n = P(z = n|X, q) = P(x_n|X, q) = \text{softmax}(s(x_n, q)) = \frac{\exp(s(x_n, q))}{\sum_{i=1}^n \exp(s(x_i, q))}$$

查询结果为 $\text{att}(\vec{X}, \vec{q}) = \sum \alpha_n \vec{x}_n$

- softmax保持了各变量的顺序
- q是查询向量，来从输入注意选择特定的项
- $s(x_n, q)$ 难的不会，就这能用简单的映射了
 - 加性函数: $s(\vec{x}, \vec{q}) = \vec{v} \tanh(\vec{W}\vec{x} + \vec{U}\vec{q})$
 - 点积函数: $s(\vec{x}, \vec{q}) = \vec{x}^T \vec{q}$
 - 归一方差缩放点积: $s(\vec{x}, \vec{q}) = \frac{1}{\sqrt{D}} \vec{x}^T \vec{q}$
 - 双线性: $s(\vec{x}, \vec{q}) = \vec{x}^T \vec{W} \vec{q}$
- 注意力的变体

1. $\text{att}(\vec{X}, \vec{q}) = \vec{x}_n, \alpha_n = \text{argmax}(\vec{\alpha}_i)$

2. key-query-value, 查询结果renew, $\text{res}' = \sum \alpha_n v_n$

3. 多头注意力机制:

$$\text{res} = \text{softmax}(s(\vec{x}, q_1))\vec{x} \oplus \text{softmax}(s(\vec{x}, q_2))\vec{x} \oplus \dots \oplus \text{softmax}(s(\vec{x}, q_n))\vec{x}$$

4. 输出序列，循环网络 c_m 用到了 $c_{1:m-1}$,

$$P(c_m | c_{1:m-1}, x_{1:N}) = \text{softmax}(s_{m,n}), s_{m,n} = \vec{v}^T \tanh(\vec{W}\vec{x}_n + \vec{U}\vec{h}_m)$$

- 序列-序列的编码/译码，

$$h_i = f_1(\vec{x}, h_{i-1}), i \in [1, T], h_{T+i} = f_2(\vec{x}, h_{T+i-1}), i \in [1, t], \text{前者为编码函数, 后者为译码函数}$$

- 这里的 $s_{m,n}$ 还是打分函数，只是查询变量更为译码向量

3. 自注意力模型，key-query-value中k,q,v不直接给出，而是由输入向量映射，

$$\vec{K}/\vec{Q}/\vec{V} = \vec{W}_{k/q/v}\vec{x}$$

4. 外部记忆-从记忆中注意目的信息，读入 $r = R(M, q_r)$ ；写入 $M = W(M, q_w, a)$ ；有端到端记忆模型和神经图灵机记忆模型

- 前者， $\vec{x} \rightarrow \vec{q}, \vec{r} = \sum \text{softmax}(s(\vec{a}, \vec{q}))\vec{v}, y = f(\vec{q} + \vec{r})$
- 后者，

$$\alpha_{t,i} = \text{softmax}(s(m_{t,n}, q_t)), r_t = \sum \alpha_{t,n} m_{t,n}, \vec{m}_{t+1,i} = \vec{m}_{t,i}(1 - \alpha_{t,i}\vec{e}_i) + \alpha_{t,i}\vec{a}_t, i \in [1, N]$$

5. 联想记忆based on 神经动力学，自联想:I/O相同空间如前馈循环net，异联想:I/O不同空间如ML方法

1. Hopfield视为循环网络，互连神经元更新公式: $s_{i,t+1} = f(\sum_{j=1}^n w_{ij}s_{j,t} + b_i)$

2. 能量 $E = -\frac{1}{2} \sum_{i,j} W_{i,j}s_i s_j - \sum_i k_i s_i$ ，最低点为稳定点，管辖区=陷入区间

3. Hebbian Rule: $W_{i,j} = -\frac{1}{N} \sum_{i=1}^N s_i^{(n)} s_j^{(n)}$, 神经元同时激活越多，连接、权重越强

- 稳定点的数量为存储容量，支持k连接的Hopfield存储容量为 $O(M^{k-1})$ ，权重可作为短期记忆