Assignment (4) Solution

24789: Intermediate Deep Learning

Theory Exercises (10 points)

PROBLEM 1

Theory [10 points]

a)

Solution:

4/5/23, 3:03 PM theory_part

In [5]:
```
import torch
import numpy as np
import torch.nn.functional as F
```

In [6]:
```
def simple_dot_product_attention(q, k, v):
    attn_logits = torch.matmul(q, k.transpose(-2, -1))
    attn_logits = attn_logits
    attention = F.softmax(attn_logits, dim=-1)
    values = torch.matmul(attention, v)
    return values
```

In [7]:
```
x = np.array([[0.8, -0.2],
              [-0.5, 0.5],
              [0, 0.5]])

w_q = np.array([[0.5, 0.5],
                [0., 1.]])

W_k = np.array([[1, 0],
                [-0.5, 0.5]])

W_v = np.array([[1, 0],
                [0, 1]])

q = torch.tensor(x @ w_q, dtype=torch.float32)
k = torch.tensor(x @ W_k, dtype=torch.float32)
v = torch.tensor(x @ W_v, dtype=torch.float32)

simple_dot_product_attention(q, k, v)
```

Out[7]:
```
tensor([[ 0.2343,  0.1863],
        [-0.0059,  0.3294],
        [ 0.0604,  0.2931]])
```

In [ ]:

localhost:8888/lab/tree/theory_part.ipynb 1/1

b) Solution: Since all elements are independent, we only need to calculate the variance of one product $q_i^l k_j^l$. Let

$X = q_i^l, Y = k_j^l$, X, Y are independent random variables.

$$\begin{aligned} Var(XY) &= E(X^2Y^2) - (E(XY))^2 \\ &= E(X^2)E(Y^2) - (E(X)E(Y))^2 \\ &= E(X^2)E(Y^2) - 0 \quad \text{(E(X)=E(Y)=0 because of standard Gaussian)} \\ &= \sigma^4 \end{aligned}$$

Hence:

$$Var(\sum_{l=0}^{d-1} q_i^l k_j^l) = \sum_{l=0}^{d-1} Var(q_i^l k_j^l) = d\sigma^4$$