# Homework - 4

1 1. softmax function is defined as $p(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$

Given values are

cat = 0.5, tiger = 0.8, human face = -3, lion = 0.6

$e^{0.5} = 1.64875$, $e^{0.8} = 2.22554$, $e^{-3} = 0.04978$,

$e^{0.6} = 1.82212$

Sum = 5.74617

$p(cat) = \frac{1.64875}{5.74617} = 0.28692$

$p(tiger) = \frac{2.22554}{5.74617} = 0.38730$

$p(human\ face) = \frac{0.04978}{5.74617} = 0.00866$

$p(lion) = \frac{1.82212}{5.74617} = 0.31710$

$$\text{probability Vector} = \begin{bmatrix} 0.28692 \\ 0.38730 \\ 0.00866 \\ 0.31710 \end{bmatrix}.$$

2. Most likely outcome is tiger with 0.38730
3. Least likely outcome is human face with 0.00866
4. cross Entropy function $H(p,q) = -\sum_i p(y_i) \log(q(y_i))$

$\Rightarrow H(p,q) = \ln(1/0.31710)$

$\Rightarrow 1.14853$.

2. 1. Recursive formula for $x_t$:

$$x_t = \beta x_{t-1} + (1-\beta) x_t$$

Bias - corrected Estimate:

$$\hat{x}_t^* = \frac{x_t}{1-\beta t}$$

$\beta = 0 \Rightarrow x_t = 0 \cdot x_{t-1} + (1-0) x_t$

$x_t = x_t$.

$\beta = 0 \Rightarrow \hat{x}_t = \frac{x_t}{1-0^t}$

since $0^t = 0$ for $t > 0$, we have

$\hat{x}_t = \frac{x_t}{1-0} = x_t$

$\therefore$ If $\beta = 0 \Rightarrow \hat{x}_t = x_t$ for all $t$

2. The recursive formula for $x_t$ is

$x_t = \beta x_{t-1} + (1-\beta) x_t$

for $t=1$ $\quad x_1 = (1-\beta) x_t + \beta \cdot 0$

$t=2$ $\quad x_2 = \beta x_1 + (1-\beta) x_2$

$\Rightarrow \quad x_t = \beta x_{t-1} + (1-\beta) x_t$

After $t$ iterations $x_t$ can be expressed as

$x_t = \sum_{i=1}^{t} \beta^{t-i} x_i$

Bias corrected $\Rightarrow \hat{x}_t = \frac{x_t}{1-\beta^t}$

Substituting $x_t \Rightarrow \hat{x}_t = \frac{\sum_{i=1}^{t} \beta^{t-i} x_i}{1-\beta^t}$

$\beta \to 1 \Rightarrow \hat{x}_t \approx \frac{\sum_{i=1}^{t} x_i}{1-\beta^t}$

$\Rightarrow \hat{x}_t \approx \frac{\sum_{i=1}^{t} x_i}{t}$

So if $\beta \to 1 \Rightarrow \hat{x}_t \to \frac{1}{t} \sum_{i=1}^{t} x_i$

3. case 1 $\rightarrow$ probability distribution

$\Rightarrow (1, -2, 3, -4)$

$e^1 = 2.71828$, $e^{-2} = 0.13533$, $e^3 = 20.0855$, $e^{-4} = 0.01831$

sum $= 22.9575$

$p(A) = \frac{2.71828}{22.9575} = 0.11840$

$p(B) = \frac{0.13533}{22.9575} = 0.00589$

$P(C) = \frac{20.0855}{22.9575} = 0.87490$

$P(D) = \frac{0.01831}{22.9575} = 0.00079$

case 2 → probability distribution

$(1, 2, -3, 0)$

$e^1 = 2.71828, e^2 = 7.38906, e^{-3} = 0.00446, e^0 = 1$

Sum = 11.1571

$P(A) = \frac{2.71828}{11.1571} = 0.24363$

$P(B) = \frac{7.38906}{11.1571} = 0.66227$

$P(C) = \frac{0.00446}{11.1571} = 0.00446$

$P(D) = \frac{1}{11.1571} = 0.08962$

1. calculate cross entropy for case 1 and case 4

$\Rightarrow H(P, Q) = -\sum_i P_i \log(Q_i) \quad \log \Rightarrow \ln$

$H(P_1, P_4) = -[0.11840 \log(0.25) + 0.00589 \log(0.25) +$
$0.87490 \log(0.25) + 0.00079(0.25)$

$= \cancel{0.} \cancel{1.3869}\ 5.24224$

$H(P_2, P_4) = -[0.24363 \log(0.25) + 0.66227 \log(0.25)$
$+ 0.00446 \log(0.25) + 0.08962 \log(0.25)$

$= \cancel{1.3863}\ 3.47989$

$H(P_3, P_4) = -[1 \log(0.25) + 0 \log(0.25) + 0 \log(0.25)$
$+ 0 \log(0.25)]$

$= \cancel{1.3863}\ \infty$

$H(P_4, P_4) = 2.$

So, the case 2 is most similar to 4.

Entropy
2. calculate for cases

$\Rightarrow$ case 1 = 3.0782

case 2 = 2.0372

case 3 = 0

case 4 = 2

So, case 4 is most similar to 3.

4. 1. when p=0 no weights are dropped, all neutrons participate in every training.

2. when p=1 all weights are dropped, leading to no learning in that layer.

3. In dropout during training, a fraction of weights are temporarily set to zero. After training is completed, dropout is no longer applied. Therefore, none of the weights are expected to have value 0.

4. Since dropout works by scaling during the testing phase, the weights should be scaled by 1-P during training.

$1 - 0.2 = 0.8$

$\Rightarrow 3 \times 0.8 = 2.4$