HW Solution

CS 512: Data Mining, Fall 2022 Version: **1.0**

Submitted by:

- **Yifei Liu**: **yifeil6**

1

Solution:

Holer

2

Solution:

2.A. Numerator is the number of samples o' "near" the sample o, denominator is the number of samples in D, thus LHS is the ratio of samples from D that is close to o.

Therefore this ratio is less than π $\Leftrightarrow$ few samples from D is close to o $\Leftrightarrow$ o is a distance-based outlier.

2.B.

$$\begin{aligned}
&\frac{\|o'|dist(o,o') \le r\|}{\|D\|} \le \pi \\
&\Leftrightarrow \|o'|dist(o,o') \le r\| \le \|D\|\pi \\
&\Leftrightarrow \|o'|dist(o,o') \le r\| < \lceil \pi\|D\| \rceil \\
&\Leftrightarrow \|o'|dist(o,o') > r\| > \lceil \pi\|D\| \rceil \\
&\Leftrightarrow \|o'|dist(o,o') > r\| > k
\end{aligned}$$

Therefore if $dist(o, o_k) > r \forall o_k$, then $\|o'|dist(o,o') > r\| > k$, thus o is an outlier. ■

2.C. Denote $f(o) = \frac{\|o'|dist(o,o') \le r\|}{\|D\|}$, then
$f(-4.5) = 0.3 > \pi$, -4.5 is not an outlier.
$f(-4) = 0.3 > \pi$, -4 is not an outlier.
$f(-3) = 0.3 > \pi$, -3 is not an outlier.
$f(-2.5) = 0.3 > \pi$, -2.5 is not an outlier.
$f(3) = 0.4 > \pi$, 3 is not an outlier.
$f(3.5) = 0.4 > \pi$, 3.5 is not an outlier.
$f(4) = 0.4 > \pi$, 4 is not an outlier.
$f(4.5) = 0.4 > \pi$, 4.5 is not an outlier.
$f(5) = 0.4 > \pi$, 5 is not an outlier.

$f(0) = 0.0 \le \pi$, 0 is an outlier.

3

Solution:

Holer

4

Solution:

4.A. a: local maximal; b: saddle point; c: cliff; d: local minimal

4.B. We have $\frac{\partial O}{\partial I} = O(1-O)$

4.B.i.

$$\begin{aligned}\delta_k &= \frac{\partial L_k}{\partial I_k} \\ &= -(T - O_k)\frac{\partial O_k}{\partial I_k} \\ &= O_k(1-O_k)(O_k - T)\blacksquare\end{aligned}$$

4.B.ii.

$$\begin{aligned}\delta_i &= \frac{\partial L_i}{\partial I_i} \\ &= \frac{\partial(\frac{1}{m}\sum_j^m w_{ij}L_j)}{\partial I_i} \\ &= \frac{\partial O_i}{\partial I_i}\sum_j^m w_{ij}\delta_j \\ &= O_k(1-O_k)\sum_j^m w_{ij}\delta_j\blacksquare\end{aligned}$$

4.B.iii.

$$\begin{aligned}delta'_k &= \frac{\partial L'_k}{\partial I_k} \\ &= \frac{\partial(-T\log O_k - (1-T)\log(1-O_k))}{\partial I_k} \\ &= \frac{\partial O_k}{\partial I_k}\frac{T - O_k}{O_k(1-O_k)} \\ &= O_k(1-O_k)\frac{T-O_k}{O_k(1-O_k)} = T - O_k\blacksquare\end{aligned}$$

4.B.iv. When output unit is saturated, we have MSE $= O_k(1-O_k)(O_k - T) \approx 0$ as well, where as for Cross-Entropy loss does not have this issue.

5

Solution:

5.A. Conv with K_1: [[0, 2, 1, 0], [1, 1, 0, 2], [2, 0, 2, 1], [1, 2, 0, 1]].
Conv with K_2: [[2, 1, 2, 2], [2, 1, 1, 2], [1, 2, 1, 3], [2, 1, 1, 2]].

5.B. Avg Pooling on feature map with K_1: [[1.0, 0.75], [1.25, 1.0]].
Avg Pooling on feature map with K_2: [[1.5, 1.75], [1.5, 1.75]].

5.C. The output shape is $E \times E \times D$, where $E = \lfloor \frac{N-K}{S} \rfloor + 1$.

5.D. $h_1 = Wh_0 + Ux_1 = [2, 2]$, $h2 = Wh_1 + Ux_2 = [5, 3]$, $\hat{y_2} = Vh_2 = [8, 5]$

6

Solution:

Holer

7

Solution:

Holer

8

Solution:

Holer

9

Solution:

Holer