Lecture 9

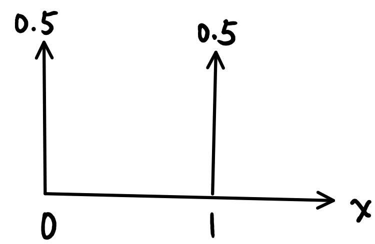
Cross Entropy

In Matlab
crossentropy
Neural network performance given targets and outputs.

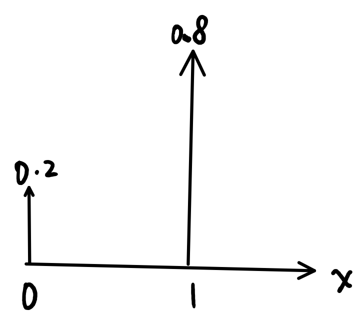
Cross Entropy Loss for classification tasks
Description:
The cross-entropy operation computes the cross-entropy loss between network prediction and target values for single-label and multi-label classification tasks.

A: True Distribution

Example :



B: "Unnatural" Distribution



Cross Entropy:

$$\boxed{H_c(A,B) = -\sum_{x\in X} P_A(x)\cdot \log_2 P_B(x)}$$

In comparison:

$$H(A) = -\sum_{x\in X} P_A(x)\cdot \log_2 P_A(x)$$

Thus,

$$H_c(A,B) = -\sum_{x\in X} P_A(x)\cdot \log_2 \underbrace{P_B(x)}_{\frac{P_A(x)\,\cdot\, P_B(x)}{P_A(x)}} = H(A) + \underbrace{\sum_{x\in X} P_A(x)\cdot \log_2 \frac{P_A(x)}{P_B(x)}}_{D_{KL}(A\|B)}$$

In summary,

$$\boxed{H_c(A,B) = H(A) + D_{KL}(A\|B)}$$

Hc(A,B): Cross Entropy -- Average number of bits needed to identify an event drawn from the alphabet if a coding scheme is "optimized" from a "unnatural" distribution B, rather than the "true" distribution A.

Go back to the numerical example:
Distribution A: PA(0) = 0.5, PA(1) = 0.5;
Distribution B: PB(0) = 0.2, PB(1) = 0.8;

Cross Entropy: $$H_c(A,B) = -\sum_{x \in X} P_A(x) \cdot \log_2 P_B(x)$$

$$= -0.5\log_2 0.2 - 0.5\log_2 0.8 = 1.3219 \text{ bits}$$

```
>> -0.5*log2(0.2) - 0.5*log2(0.8)
ans =
   1.3219
```

Alternatively,

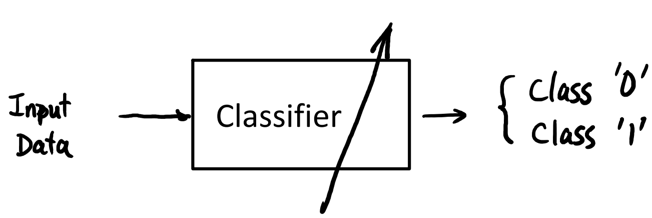
$$H_c(A,B) = H(A) + D_{KL}(A\|B)$$

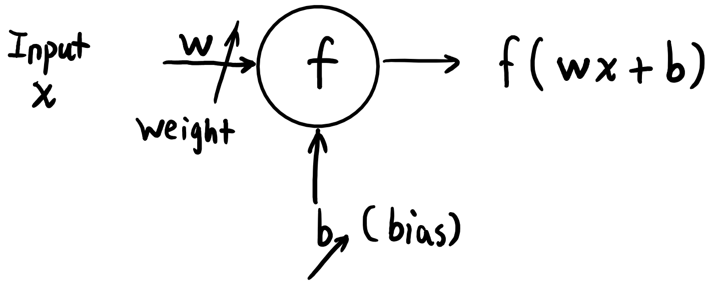
$$= 1 + \underbrace{0.5\log_2 \frac{0.5}{0.2} + 0.5\log_2 \frac{0.5}{0.8}}_{0.3219 \text{ bit}}$$

$$= 1.3219 \text{ bits}$$

```
>> 0.5*log2(0.5/0.2) + 0.5*log2(0.5/0.8)
ans =
   0.3219
```

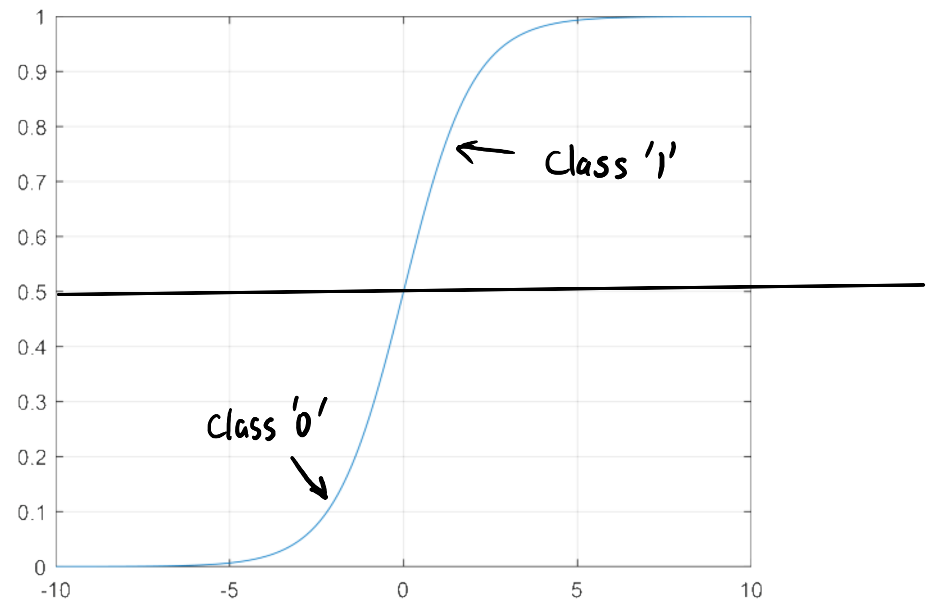
- Loss Function in Machine Learning and Optimization





Example of the function

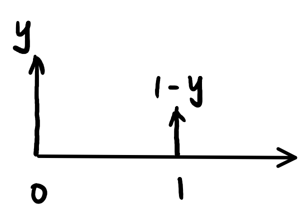
$$f(z) = \frac{1}{1+e^{-z}}$$



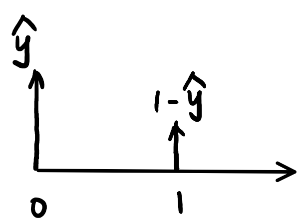
A: — True Labels

B: — Predicted labels decided by the classifier

A:



B:



Cross Entropy: $H_c(A, B) = -y \log_2 \hat{y} - (1-y) \log_2 (1-\hat{y})$.

- Entropy Estimation
  Example: Alphabet = {1, 2, 3}.

Sequence : 1, 2, 1, 2, 3, 3, 3, 3, 1, 2, 3, 3, 3, 3, 1, 2, 3, 3, 1, 2

$P(1) = \frac{5}{20} = \frac{1}{4}$, $P(2) = \frac{5}{20} = \frac{1}{4}$, $P(3) = \frac{10}{20} = \frac{1}{2}$.

First-order Entropy: $H(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}) = 1.5$ bits/symbol

Consider block of two symbols:

Sequence : 1, 2, 1, 2, 3, 3, 3, 3, 1, 2, 3, 3, 3, 3, 1, 2, 3, 3, 1, 2

....

Joint Distribution:

$P(1,1) = \frac{0}{19} = 0$ $\quad$ $P(1,2) = \frac{5}{19}$ $\quad$ $P(1,3) = \frac{0}{19}$

$P(2,1) = \ldots$ $\quad$ $P(2,2) = \ldots$ $\quad$ $P(2,3) = \ldots$

$P(3,1) = \ldots$ $\quad$ $P(3,2) = \ldots$ $\quad$ $P(3,3) = \frac{7}{19}$

Conditional Prob's:

Use another example : Consider binary sequence : w w b b b b b w w

Alphabet = $\{b, w\}$

$P(w|w) = \frac{2}{3}$, $P(b|w) = \frac{1}{3}$, $P(w|b) = \frac{1}{5}$, $P(b|b) = \frac{4}{5}$

next symbol $\quad$ current symbol