Note: This is the summary note from Udacity Introduction to Deep Learning with PyTorch

Softmax

- Problem: multi-classification
- Solution: satisfy probability, here are two criterias
  - Total sum of the scores need to be one
  - The higher the score the higher the probability
- How?
  - Use Softmax function: Turns $\frac{\text{score}}{sum\ of\ all\ scores}$ to positive number with exponential
  - Assume we have linear function score: $z_1 + z_2 + .. + z_n$
    So the Softmax function for probability of class i is
    $P(class\ i) = \frac{e^{z_i}}{e^{z_1} + e^{z_2} + .. + e^{z_n}}$

One Hot Encoding

- Input data will not always looks like number
- Solution: turn data to number, One Hot Encoding
- How:
  - To avoid dependencies between data we create table of data in which 1 represent correct data input while 0 for the rest of incorrect data

Maximum Likelihood

- Pick the model that give existing labels the highest probability
- The best model more likely be give highest probability to the event that happens to the true label (ground-truth)
- Maximize probability -> get best model
- How:
    - Start with bad model
    - Calculate probability of each point
    - P = Multiple (Product) the probability of each point
    - Find way to maximize P

Maximizing Probabilities

- Probability is important
- Maximize probability --> minimized the error function
- Problem with product
    - Hard when have thousands data point
    - The number is small between 0 to 1
    - If change 1 data point => the whole product change drastically
- Solution
    - Do sum by using log

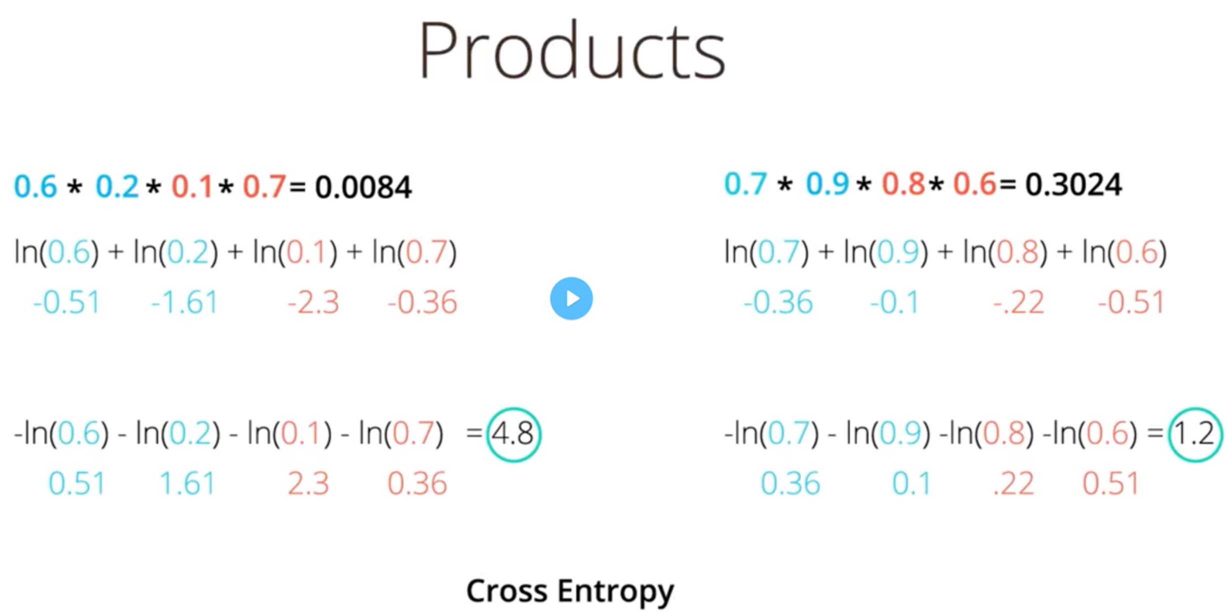
Cross Entropy

- Resolve product problem above by using log

$$log(ab) = \log(a) + \log(b)\ ;$$

Note: if log(x) provide negative value, we need to take the

$$log(1) = \log(0)$$



- The sum of negative of log of probability -> Cross Entropy
- The bad model give high cross entropy
- The good model give low cross entropy
- **Goal is to minimize the cross entropy**
- Connection between probability and error function:
  - Events
  - Probability
  - Cross Entropy

- How often the events happens based on probability
  - if it's likely --> small cross entropy
  - If it's unlikely --> large cross entropy
- Formula :

$$Cross\ Entropy = -\sum_{i=1}^{m} y_i ln(p_i) + (1 - y_i)\, ln(1 - p_i)$$

Y label
P = probability that event occur

Multi class cross entropy

- Problem: what if we have multiple class?
- $Cross\ Entropy = -\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} ln\left(p_{ij}\right)$
  - m is number of classes

==
Mini summary:

- Softmax function used for multi-class
- One Hot encoding is used when input data is not numeric
- Maximized probability: pick the model that give existing labels the highest probability
- Cross Entropy is connection between probability and error function.
- How often the events happens based on probability
  - if it's likely --> small cross entropy
  - If it's unlikely --> large cross entropy

==