

### Quiz 1-1: Back Propagation (20%)

Handwritten mathematical derivations for the cross-entropy loss function and its derivatives:

$$CE(y, \hat{y}) = - \sum y \log \hat{y}$$
$$\hat{y} = \frac{e^{\theta}}{\sum e^{\theta}}$$
$$CE(y, \hat{y}) = - \sum y \log \left( \frac{e^{\theta}}{\sum e^{\theta}} \right)$$
$$= - \sum y \left( \theta - \log(\sum e^{\theta}) \right)$$
$$\frac{\partial}{\partial \theta} CE(y, \hat{y}) = - \sum y = -y_x$$
$$\theta = W_2 h + b_2$$
$$CE(y, \hat{y}) = - \sum y (W_2 h + b_2 - \log(\sum e^{\theta}))$$
$$= - \sum y W_2 h - \sum y b_2 + \sum y \log(\sum e^{\theta})$$
$$\frac{\partial}{\partial W_2} CE(y, \hat{y}) = - \sum y h = y h_x$$

### Quiz 1-2: Common Senses (30%)

Note: you can answer in Chinese or English.

Note: your answers can be as simplest as possible

- Why we use F1-score instead of precision(accuracy)?  
F1-score includes precision and recall accuracy.  
Precision represents that the accuracy when the classifier says one case is positive and the case is really positive.  
Recall represents that the accuracy when one case is positive and the classifier does say it is positive.  
If we use precision as our measure, we may miss a lots of positive cases (low recall).
- Why don't we use binary classification function as the activation function in neural networks?
- What is the bias and variance of a machine learning algorithm?  
Bias: How well when we train a model.  
Variance: How well when we test a model.
- When training a single tree in random forest, the tree is never pruned, why?

- What is one-hot encoding?  
There is only one bit is 1 and the others are 0.
- How to prevent overfitting in neural networks? (write down anything you know)
  1. Stop training earlier.
  2. Delete some neurons (simpler model).
  3. Use more data to train.