Notes DL01 Loss function

Slide 20 Handling Multiple Classes

$$P_i = \frac{\exp(y_i)}{\sum_{j=1}^{C} \exp(y_j)}$$

* $P_i \in [0,1]$

$\sum P_i = 1$

the bigger the score is, the higher the possibility is.

$y_1, \dots y_C$ are the C outputs of our network

q: we construct a one-hot encoding q of the class label t.

$H(q,p)$: Cross entropy,

$$H(q,p) = -\sum_{i=1}^{C} q_i \log p_i$$

$$\begin{aligned} l(y,t) = H(q,p) &= -\sum_{i=1}^{C} q_i \log p_i \\ &= -\log p_t \\ &= -\log \frac{\exp(y_t)}{\sum_{j=1}^{C} \exp(y_j)} \\ &= \log \sum_{j=1}^{C} \exp(y_j) - y_t \end{aligned}$$

★ When the loss is small, the score y_t of the correct class must be higher than the score of all other classes.

Example:

assume three classes: $y = (y_1, y_2, y_3)$

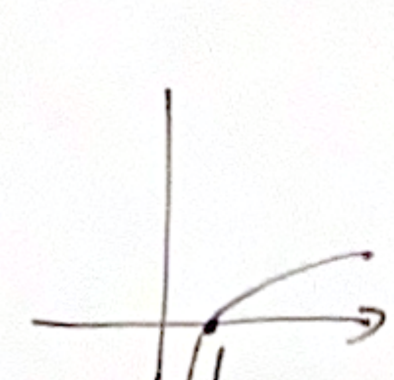
the true label is $t=1$

Case A: Wrong classes have higher scores.

$y = (1, 10, 8)$

$$P_1 = \frac{e^1}{e^1 + e^{10} + e^8} \approx 0$$

$$l = -\log P_1 = -\log(0) = +\infty$$



Case B:

$y = (6, 3, 2)$

$$P_1 = \frac{e^6}{e^6 + e^3 + e^2} \approx 0.94$$

$$l = -\log(0.94) \approx 0.06$$