Pseudo Label

分为两个阶段: pre-training 和 fine-tuning。pre-training就是用labeled data训练出一个网络，fine-tuning就是生成pseudo label。

等同于最小化无标签数据的类别概率的条件熵，减少class overlap。本篇就是讲述了为什么pseudo label有用，翻译一下就是加入置信度高的unlabel数据让结果更加集中。

Loss function:

$$L = \frac{1}{n} \sum_{m=1}^{n} \sum_{i=1}^{C} L(y_i^m, f_i^m) + \alpha(t) \frac{1}{n'} \sum_{m=1}^{n'} \sum_{i=1}^{C} L(y_i'^m, f_i'^m) \qquad (1)$$
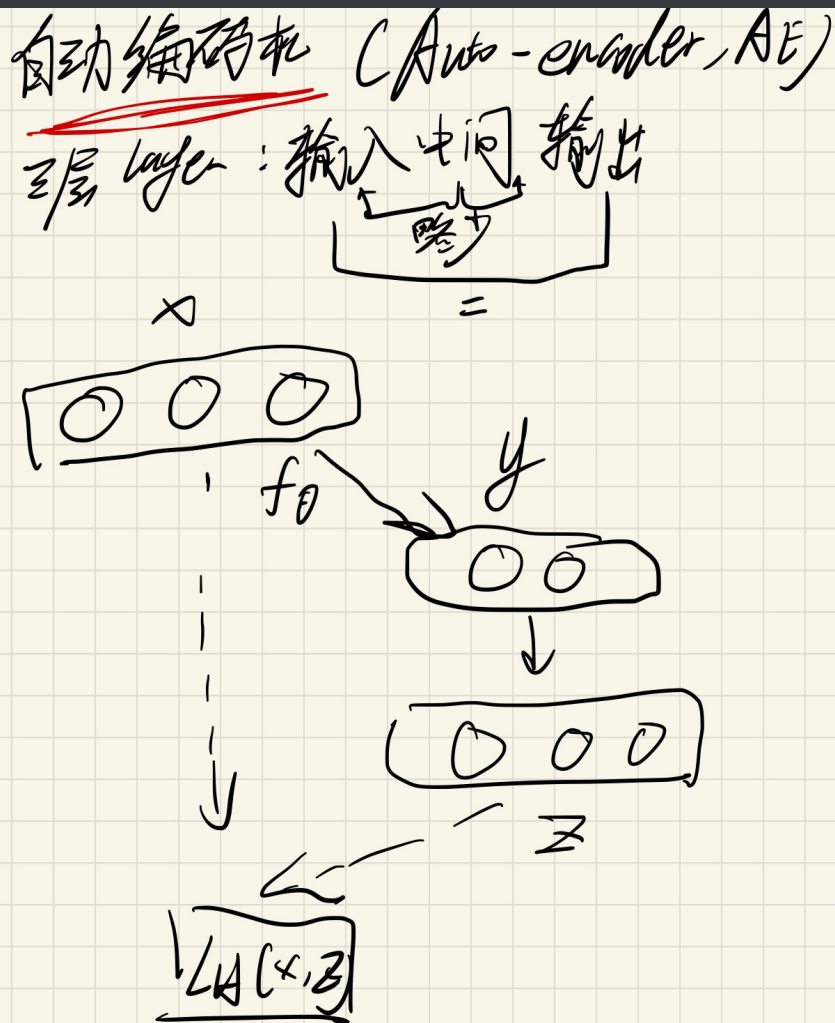
L为cross-entropy loss function。

α:

$$M_1 = \begin{cases} 0 & t < T_1 \\ \frac{t-T_1}{T_2-T_1} \alpha_f & T_1 \le t < T_2 \\ \alpha_f & T_2 \le t \end{cases}$$
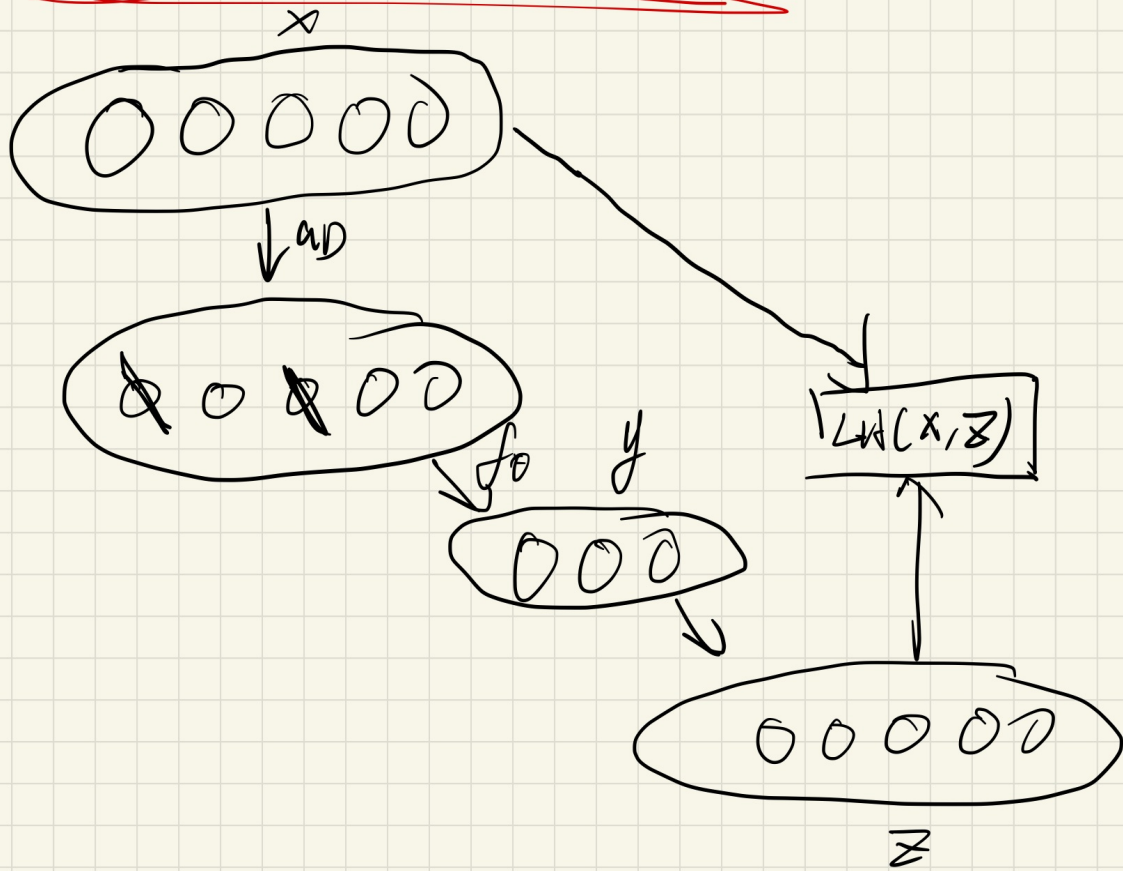
Class overlap

class overlap是一个用来检测模型性能的变量，class overlap越大，模型性能越差。

Denosing Auto-Encoder

一种降维手段。

自动编码机 (Auto-encoder, AE)

三层 layer：输入 中间 输出

略去

$=$

X

f_θ

y

Z

$LH(X,Z)$

降噪自动编码器 (DDAE)

x

00000

$\downarrow aD$

$\cancel{0} 0 \cancel{0} 00$

f_θ

y

$LH(x, \vec{z})$

000

00000

z

Dropout