

\$notebook\_name\$

\$number\_of\_pages\$

\$new\_page\$

\$notebook\_created\$

\$notebook\_modified\$

# Autoencoder-based Graph Construction for SSL

Problem formulation:

$$L = \{(x_i, y_i)\}_{i=1}^m, \quad U = \{x_i\}_{i=m+1}^n$$

$$x_i \in X, \quad y_i \in Y = \{1, 2, \dots, c\}$$

loss 由 label data 的 supervised loss 与 unlabeled data 的 regularization term 组成.

regularization terms: consistency loss + feature matching loss

Consistency loss:

$$L_c(\theta) = \sum_{i=1}^n L_c(f(x_i; \theta, \xi), \tilde{f}(x_i; \theta', \xi'))$$

其中,  $f(x_i; \theta, \xi)$  为模型预测结果,  $\xi$  为施加于  $x_i$  上的随机扰动

$\tilde{f}(x_i; \theta', \xi')$  为 teacher model,

$L_c(\cdot)$  为距离函数

Feature matching loss:

$$L_g(\theta, W) = \sum_{x_i, x_j \in L \cup U} L_g(h(x_i; \theta), h(x_j; \theta), W_{ij})$$

其中,  $h: X \rightarrow \mathbb{R}^p$  为输入空间到低维特征空间的映射.

即... 相同矩阵  $W$  的  $(i, j)$  项 (同类为 1 异类为 0)

$W_{ij}$ : 相似及图元素  $W_{ij}$  的权重 (通常,  $W_{ij} \in [0, 1]$ )

$l_g(\cdot, \cdot)$ : 取决于  $W_{ij}$  的一种距离度量

$$l_g = \begin{cases} \|h(x_i) - h(x_j)\|^2, & \text{if } W_{ij} = 1 \\ \max(0, \text{margin} - \|h(x_i) - h(x_j)\|)^2, & \text{if } W_{ij} = 0 \end{cases}$$

margin 为预定义的正数.

使用 matrix completion 的方法构建  $W$ .

SBL Framework.

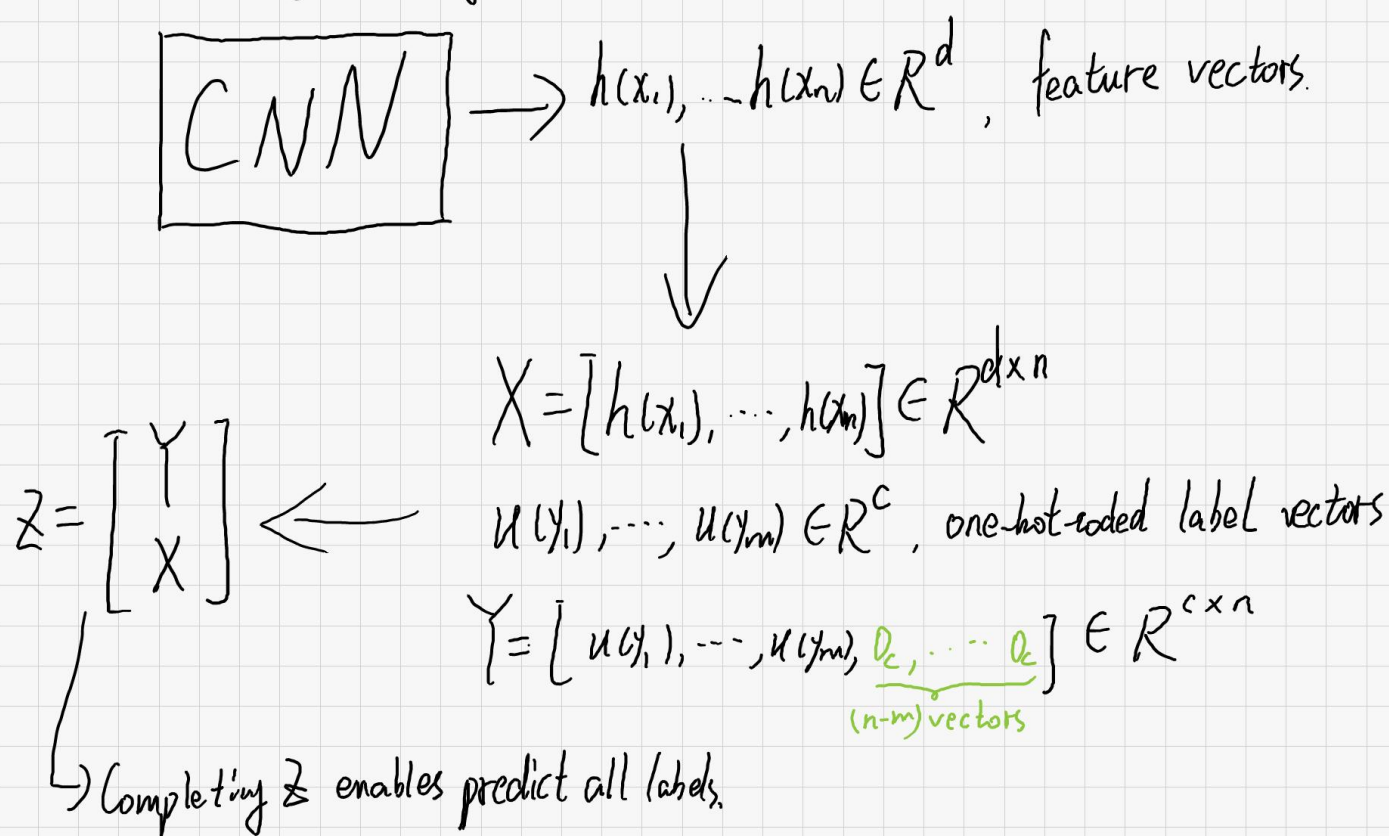
$$L_s(\theta) + \lambda_c L_c(\theta) + \lambda_g L_g(\theta) \quad (5)$$

# Autoencoder-based Graph Construction for SSL

## Our approach

为减少计算开销, 使用神经网络进行 matrix completion 工作  
使用 auto-encoder 计算参数, 记为  $\phi$ , 使用 **autoencoder loss** 计算开销及 Graph estimate.

### 4.1. Learning the graph with autoencoder



input dropout: 以一定概率将 label-presence column 向量归 0, 即  $z_j = \begin{bmatrix} u(y_j) \\ h(x_j) \end{bmatrix} \rightarrow \bar{z}_j \begin{bmatrix} 0_c \\ h(x_j) \end{bmatrix}$

最终得到  $\bar{Z} = \begin{bmatrix} \bar{Y} \\ X \end{bmatrix}$  (7)

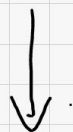
autoencoder loss:

$$\mathcal{L}_{\text{AE}}(\phi) = \sum_i (\bar{Y}_{:,i} - \hat{\bar{Y}}_{:,i})^2 + \mu \|\bar{X} - \hat{\bar{X}}\|^2 \quad (8)$$

$$L_{AE}(\theta, \phi) = \sum_{(x, y) \in \Omega_T} \mu(t) \cdot \|g_\phi(x) - y\|_F$$

其中,  $(\hat{y}, \tilde{y})$  表示 autoencoder  $g_\phi(\tilde{x})$  的输出.

$\mu(t)$  为单调增函数.



记  $\tilde{u}(y_i)$  为 predict label vector,  $\tilde{y}_i = \operatorname{argmax}_k [\tilde{u}(y_i)]_k$

因此, similarity graph  $W$  为:

$$W_{ij} = \begin{cases} 1, & \text{if } \tilde{y}_i = \tilde{y}_j \\ 0, & \text{if } \tilde{y}_i \neq \tilde{y}_j \end{cases}$$

## 4.2 Simultaneous training

$$L_S(\theta) + w(t) \cdot (L_c(\theta) + \lambda \cdot L_g(\theta, W)) + L_{AE}(\theta, \phi) \quad (10)$$

文献[18][24]

Experiments

数据集: MNIST, SVHN, CIFAR-10