Ⅰ label noise 概述

harmful.
- training: overfit
- Validation: network is nonconvex.

ubiquitous.

DN 引入 noise 后 泛化能力大大降低

方法1.

extract confident example.

Memorization effect: prefer major pattern.

During early training stage, examples with small loss are likely to be true.

PES RSA

方法2

Semi-supervised.

DivideMix

方法3.

Label noise learning.

噪声建模. class conditional

$$p(x|\tilde{Y}=j) = \sum_{i=1}^{C} \alpha_{ij}\, p(\vec{x}|Y=i).$$

class posterior

$$p(\tilde{Y}=i|x) = \sum_{i=1}^{C} \beta_{ij}\, p(Y=i|\vec{x}).$$

1.

Label Noise Algorithm
- Stastically inconsistent algorithm (heuristic).
  - select clean label
  - re-weighting
  - correcting labels.
- SCA. → design classifier-consistent algorithms

  s.t. converge to classifier on clean domain!!!

  Statistically consistent algorithm.

SCA:

clean label transition matrix: $T(x) = p(\tilde{Y}|Y, X=\vec{x})$.

basic math: clean label classifier → clean class posterior posterity $p(Y|X)$. (C). → Through learning, $f(x) := \arg\max_{Y} p(Y|X)$.

clean label distribution.

↓ 通过乘 T (T).

noisy label posterity $p(\tilde{Y}|X)$. (N) → classifier learned in real dataset.

So. $N = TC$.

Expected Risk: $R_{D,0\text{-}1}(f) = E_{(X,Y)\sim D}[1(Y \neq \mathrm{sign}(f(x)))]$

近似

Empirical Risk: $\hat{R}_{D,L}(f) = \frac{1}{n}\sum_{i=1}^{n} l(y_i, f(x_i))$.

2. T is determined by the type of NL.
- RCN. } instance-independent
- CCN. }
- IDN — instance-dependent

a. Assume T is predetermined.

① RCN

$\boxed{\text{Thm.}}$ As long as $L(f(x),+1) + L(f(x),-1) = C$.

then, $\arg\min_f R_{D,L}(f) = \arg\min_f R_{\hat{D},L}(f)$.

because,

$$R_{\hat{D},L}(f) = E_{(X,\tilde{Y})\sim\hat{D}}[L(f(X),\tilde{Y})] = (1-2\rho)R_{D,L}(f) + \rho C$$

②. CCN

$\boxed{\text{Thm.}}$ Modify l to be $\hat{l}$ that

$$E_{(X,\tilde{Y})\sim\tilde{D}}[\hat{l}(f(x),\tilde{Y})] = E_{(X,Y)\sim D}[l(f(x),Y)].$$

for binary classification, with the help of T.

unbiased estimator is derived.

③. IDN.

Importance reweighting.

forward correction.

b. how to estimate T?

$\underline{\underline{N}} = TC$ suffers from identifiability $(A = TB = (TQ)(Q^{-1}B))$.

KNOWN

so there has to be prior condition.

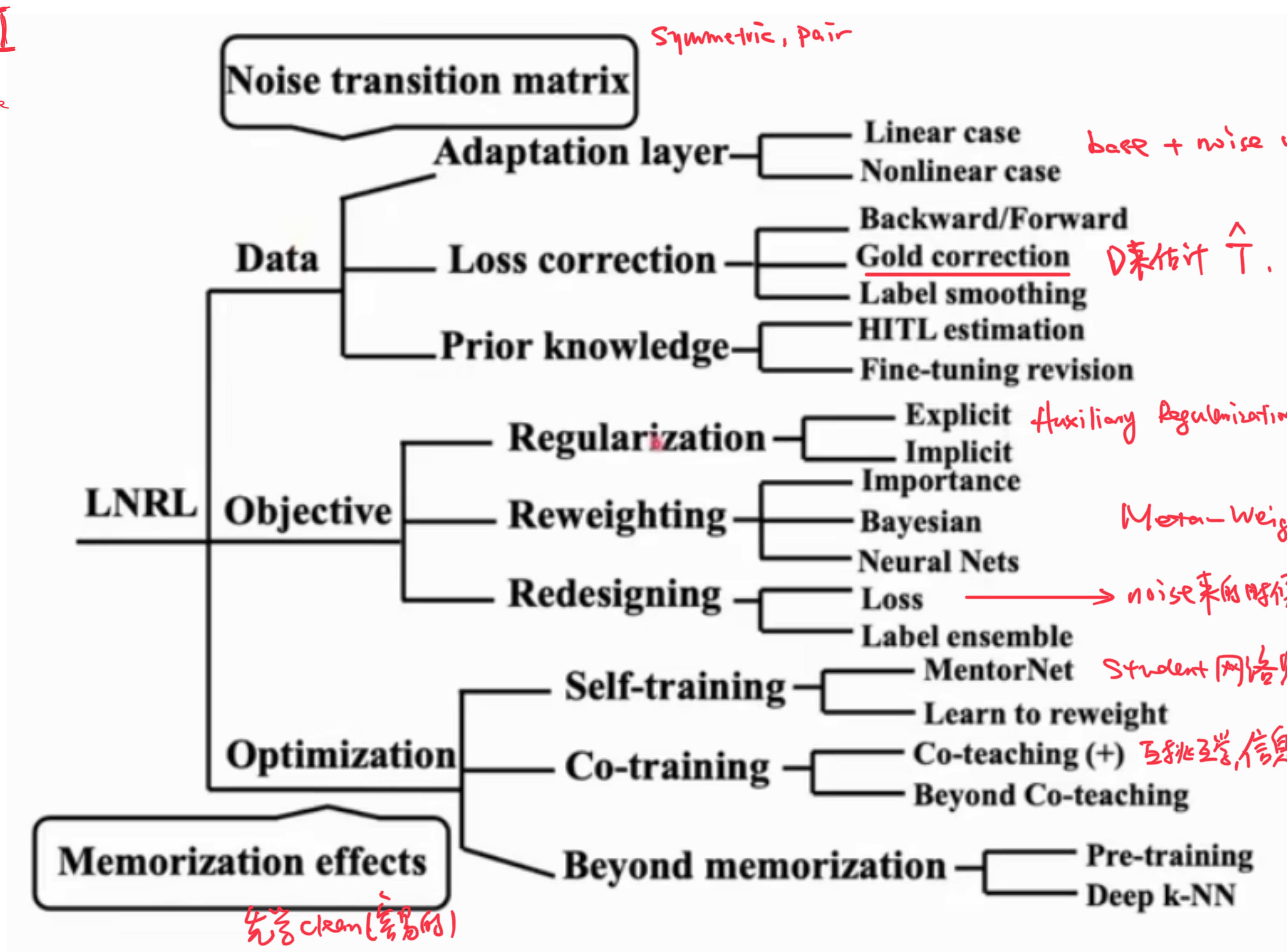
instance-independent
- ① anchor point Asum. if ap exists, $\beta_{-y,+y} = \min_{\vec{x}\in X} (\tilde{Y}=y|\vec{x})$.
- ② sufficiently scattered Asum.

instance-dependent.
- ① part-dependent.
- ...

Ⅱ

Label Noise 常用算法分类

symmetric, pair



base + noise model.

D来估计 $\hat{T}$. *Meta Loss correction.

Auxiliary Regularization: 惩罚过多样本(优先clean).

Meta-Weight-Net.

→ noise来的时候loss大,降低关注.

Student 网络只学干净, mentor 筛选干净样本.

互相学习,信息互补

先学clean(容易的)

Small-loss trick: 只学小loss的样本,其他淘汰

更难的场景: instance-dependent, OOD noise.