

2024-09-11

~~DOMAIN ADAPTATION~~

DOMAIN ADAPTATION

Typical ML problem

- Training Set $\left\{ (x_i^{train}, y_i^{train}) \right\}_{i=1}^m \sim \underline{Q}_{X,Y}$

- Goal: $f_{\theta}: X \rightarrow Y$

- Objective: $\min_{\theta} \frac{1}{m} \sum \ell (f_{\theta}(x_i^{train}), y_i^{train})$

$\theta^* = \underset{\theta}{\operatorname{argmin}}$ _____ " _____

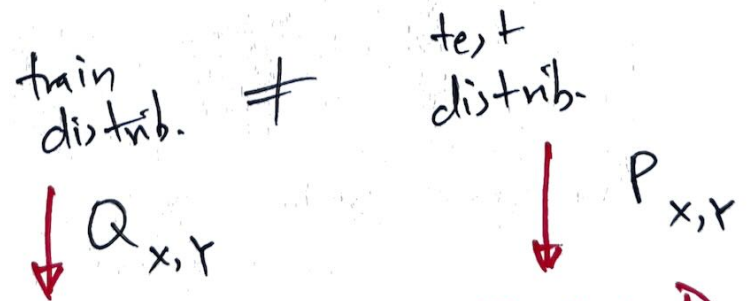
- Test Set $\left\{ (x_i^{test}, y_i^{test}) \right\}_{i=1}^n \stackrel{i.i.d.}{\sim} \underline{Q}_{X,Y}$

- Test Error $\frac{1}{n} \sum_{i=1}^n \ell (f_{\theta^*}(x_i^{test}), y_i^{test})$

GOAL: \uparrow
SMALL

- Assumption: both TRAIN, TEST $\sim Q_{X,Y}$

- in most real-world case



SOURCE DOMAIN

TARGET DOMAIN

VISION
day-time photographs in California
SP EECH British accent (ENGLISH)

night-time photo. in Maryland
Indian accent (ENGLISH)

D-A. PROBLEM SETUP

TRAINING

— for source domain, we have labeled examples

$$(X_s, Y_s) = \left\{ (X_i^s, Y_i^s) \right\}_{i=1}^m \sim Q_{X, Y}$$

— for target domain

CASE ① we only have unlabeled target samples

$$X_T = \left\{ X_i^t \right\}_{i=1}^{m_t} \sim P_X$$

" UNSUPERVISED D.A "

CASE ② = Some labeled target samples
& Some unlabeled target samples

$$X_{T_1} = \left\{ X_i^t \right\}_{i=1}^{m_t} \quad \text{and} \quad X_{T_2}, Y_{T_2} = \left\{ (X_i^t, Y_i^t) \right\}_{i=1}^{n \ll m_t}$$

" SEMI-SUPERVISED D.A "

CASE ③ NO target data

" DOMAIN GENERALIZATION "

OR

" out of distribution GENERALIZATION "

UNSUPERVISED D-A

Given: labeled samples from $Q_{x,y}$
unlabeled samples from P_x

Pick a function class
"hypothesis class" H
(eg. a neural network architecture)
with weights θ

Goal: find the best $h \in H$
s.t. target error is minimized

$$\epsilon_t(h) = \mathbb{E}_{x,y \sim P_{x,y}} [l(h(x), y)]$$

↓
target error

↓
target distrib.

ASSUMPTIONS

- ① SRC and TARGET Distributions are "similar"
- $$Q_x \stackrel{\text{"similar"}}{\sim} P_x$$

(2) COVARIATE SHIFT ASSUMPTION

two distri'b. P and Q satisfy the cov-shift assumption

if $P(y|x) = Q(y|x) \quad \forall x \in X$

the conditional label distribution does not change betⁿ SRC & TRG.

the true probability of labels does not change

(3) if I hard labelled target samples during training, the joint error should be small.

$$\begin{aligned} \mathcal{E}_{\text{joint}} &= \frac{1}{m} \sum_{i=1}^{m_s} \ell(h(x_i^s), y_i^s) \rightarrow \text{source error} \\ &+ \frac{1}{m} \sum_{i=1}^{m_t} \ell(h(x_i^t), y_i^t) \rightarrow \text{target error} \end{aligned}$$

if we have ~~access to~~ full access then we can optimize to find weights that minimize joint ~~loss~~ error.