2024-09-11

~~DOMAIN ADAPTATION~~

# DOMAIN ADAPTATION

Typical ML problem

- Training Set $\{(x_i^{train}, y_i^{train})\}_{i=1}^{m} \sim Q_{X,Y}$
- Goal : $f_\theta : X \to Y$
- Objective : $\min_\theta \frac{1}{m} \sum \ell(f_\theta(x_i^{train}), y_i^{train})$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \text{ ——— '' ——— }$$

- Test Set $\{(x_i^{test}, y_i^{test})\}_{i=1}^{n} \overset{i.i.d}{\sim} Q_{X,Y}$
- Test Error $\frac{1}{n} \sum_{i=1}^{n} \ell(f_{\theta^*}(x_i^{test}), y_i^{test})$

GOAL: SMALL

- Assumption : both TRAIN, TEST $\sim Q_{X,Y}$
- in most real-world case train distrib. $\neq$ test distrib.

train distrib. $\to Q_{X,Y}$ → SOURCE DOMAIN

test distrib. $\to P_{X,Y}$ → TARGET DOMAIN

| | SOURCE DOMAIN | TARGET DOMAIN |
|---|---|---|
| VISION | day-time photographs in California | night-time photo.. in Maryland |
| SPEECH | British accent (ENGLISH) | Indian accent (ENGLISH) |

# D.A. PROBLEM SETUP — TRAINING

- for source domain, we have labeled examples

$$(X_s, Y_s) = \{(x_i^s, y_i^s)\}_{i=1}^{m} \sim Q_{X,Y}$$

- for target domain

CASE ① we only have unlabeled target samples

$$X_T = \{x_i^t\}_{i=1}^{m_t} \sim P_X$$

" UNSUPERVISED D.A "

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

$$X_{T_1} = \{x_i^t\}_{i=1}^{m_t} \quad \text{and} \quad X_{T_2}, Y_{T_2} = \{(x_i^t, y_i^t)\}_{i=1}^{n \ll m_t}$$

" 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 $\theta$

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

$$\underbrace{\varepsilon_t(h)}_{\text{target error}} = \mathbb{E}_{x,y \sim \underbrace{P_{X,Y}}_{\text{target distrib.}}} \left[ \ell(h(x), y) \right]$$

ASSUMPTIONS:

① SRC and TARGET Distributions are "Similar"

$$Q_X \overset{\text{"similar"}}{\sim} P_X$$

## (2) COVARIATE SHIFT ASSUMPTION

two distrib. P and Q satisfy the cov.shift assumption

<u>if</u> $P(y|x) = Q(y|x) \quad \forall x \in X$

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

the true probability of labels does not change

(3) <u>If</u> I "<u>~~had~~</u>" labelled target samples during training,

the joint error should be ~~s~~ small.

$$\mathcal{E}_{joint} = \frac{1}{m} \sum_{i=1}^{m_s} \ell(h(x_i^s), y_i^s) \rightarrow \text{source error}$$

$$+ \frac{1}{m} \sum_{i=1}^{m_t} \ell(h(x_i^t), y_i^t) \rightarrow \text{target error}$$

<u>If</u> we have ~~access to~~ full access

then we <u>can</u> optimize to find weights that minimize joint ~~loss~~. error.