Summary: Why and how to correct for target/conditional shift?

- **Problem**: predicting Y from X, under P_{TX} ≠ P_{XY} and P_{TX}(X) ≠ P_{Y}(X), but it is plausible to assume:
  - **Target shift (TarS)**: P_{TX} ≠ P_{XY} and P_{TX}(X) ≠ P_{Y}(X),
  - **Conditional shift (ConS)**: P_{TX} ≠ P_{XY} and P_{TX}(X) = P_{Y}(X), and
  - **Generalized target shift (GeTarS)**: P_{TX} ≠ P_{XY} and P_{TX}(X) ≠ P_{Y}(X).
- **Current approaches**:
  - Inefficient methods to correct for ConS and GeTarS with kernel mean matching

### Possible situations for domain adaptation

- **Target shift (or prior probability shift)**: P_{TX} helps predict Y.

---

### Distribution shift correction by data transformation/reweighting

- **Problem**: Given training data D^T = (x_i,y_i)_{i=1}^m, find the regressor (e.g., KRR) or classifier (e.g., SVM) f(x) that works well on test data D^T = (x_i,y_i)_{i=1}^m.

#### Importance reweighting

- **Minimize the expected loss on test data**:

\[
R[f^*, \theta; x,y, \theta] = E_{X,Y \sim P_{TX,Y}}[f(x, \theta)y] = E_{X,Y \sim P_{TX,Y}}[P_{TX}[f(x, \theta)] y] = \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy.
\]

- **Empirical version**:

\[
\hat{R}[f^*, \theta; x,y, \theta] = \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy = \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy.
\]

- **Sample transformation and reweighting**:

\[
\hat{R}[f^*, \theta; x,y, \theta] = \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy = \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy.
\]

- **Consider** (K^*θ,ψ)^T as new training data and learn under TarS.

- **Will be used to correct for GeTarS Problem**: How to find β and/or T?

### Correction for target shift (Fig. 3)

- **Aim** to find β and P_{TX} = P_{XY} under TarS P_{TX} = P_{XY} but P_{TX} ≠ P_{XY}.

- **Richness of training data**: the support of P_{TX} contains that of P_{XY}.

- **Invertibility**: only one distribution of Y, together with P_{TX} leads to P_{XY}.

- **Kernels k (for X) and l (for Y) are characteristic**.

#### Traditionally difficult, but very convenient with kernel mean matching

- P_{XY} has a unique embedding \mu(P_{XY}) with characteristic kernels.

- Avoid explicit estimation of P_{XY}.

- Conditional embedding is an operator from \mathcal{F} to \mathcal{T} (\mathcal{F}(X) = \mathcal{Y} X \mathbb{C}_X \mathbb{C}_Y, \mathcal{X} \quad \text{and} \quad \mathcal{Y} \quad \text{are unoriented cross-}

- auto-covariance operators).

- P_{XY} = \mathcal{Y} X \mathbb{C}_X \mathbb{C}_Y.

- P_{TX} = \mathbb{K} (K + \mathbb{M})^{-1} \mathbb{T}.

#### Let P_{TX} = \beta P_{XY}.

- Find f by matching P_{XY} (corresponding to P_{Y}(X)) and P_{TX} (corresponding to P_{XY}).

#### Objective function

\[
J = \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy.
\]

- **Regularization on \{W, \beta\} for stability**.

### Real-world problems

- **Regression under TarS**

- **Covariate pair (time series Y and X):** X is the input sent by the computer, with a strong dependence.

- **TarS improves prediction performance for Y**.

- **No improvement for predicting X from Y**.

- **Remote sensing image classification**

- **Unlabeled image data set** is used on the computer side for both TarS and GeTarS.

### Location-scale generalized target shift (Fig. 4)

- **Assumption**: Both P_{TX} and P_{XY} change, but P_{XY} changes only in the location and scale, i.e., \mathbb{P}(Y | X) = \mathbb{P}(y | X) and \mathbb{P}(Y | X) = \mathbb{P}(y | X) = \mathbb{P}(y | X).

- **Identifiability**: Under certain conditions on P_{XY}(x), \mathbb{P}(Y | X) and \mathbb{P}(Y | X) uniquely recovered by reweighting and transforming training data to reproduce P_{TY} via, by minimizing

\[
\left\| \mu(P_{TY}) - \mu(P_{TX}) \right\|.
\]

- **Objective function**

\[
J = \frac{1}{m} \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy = \sum_{i=1}^m \beta_i \hat{y}_i^2(x_i, \theta) dx dy.
\]

- **Domain adaptation under Target and Conditional Shift**

- **Problem**: domain adaptation under target and conditional shift.

- **Goal**: find the regressor (e.g., KRR) or classifier (e.g., SVM)

- **Optimization**

- **Minimize the expected loss on test data**:

\[
\left\| \mu(P_{TY}) - \mu(P_{TX}) \right\|
\]

- **Domain adaptation under location-scale target shift**

- **Problem**: domain adaptation under location-scale target shift.

- **Goal**: find the regressor (e.g., KRR) or classifier (e.g., SVM)

- **Optimization**

- **Minimize the expected loss on test data**:

\[
\left\| \mu(P_{TY}) - \mu(P_{TX}) \right\|
\]

- **Conclusion**

- **TarS and GeTarS a convenient way to deal with the situation where both conditional and marginal distributions change across domains**.

- **Background (causal) information helps learning**: compact description of how distributions change.

- **Reference**