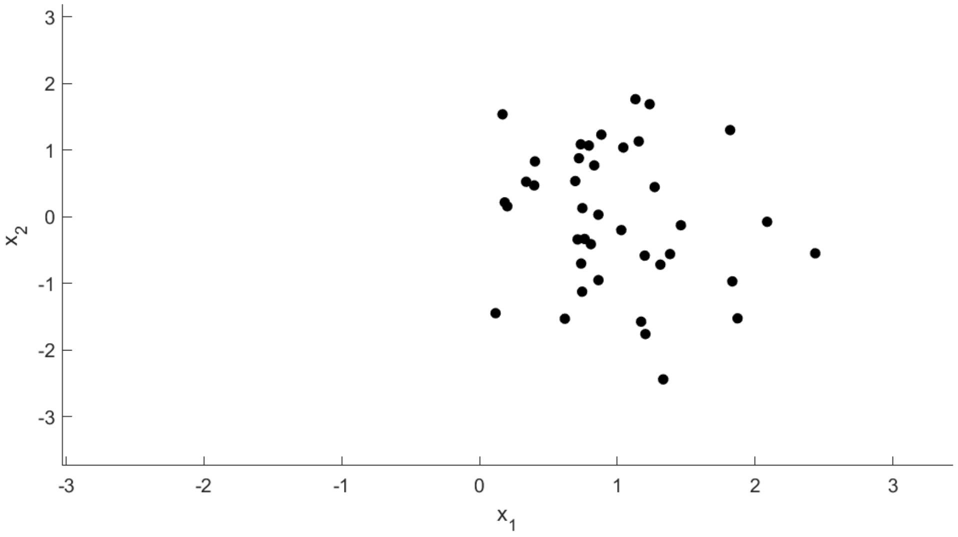
Target Robust Discriminant Analysis

Wouter Kouw & Marco Loog
IAPR International Workshops on Statistical + Structural and Syntactic Pattern Recognition (S+SSPR 2020)

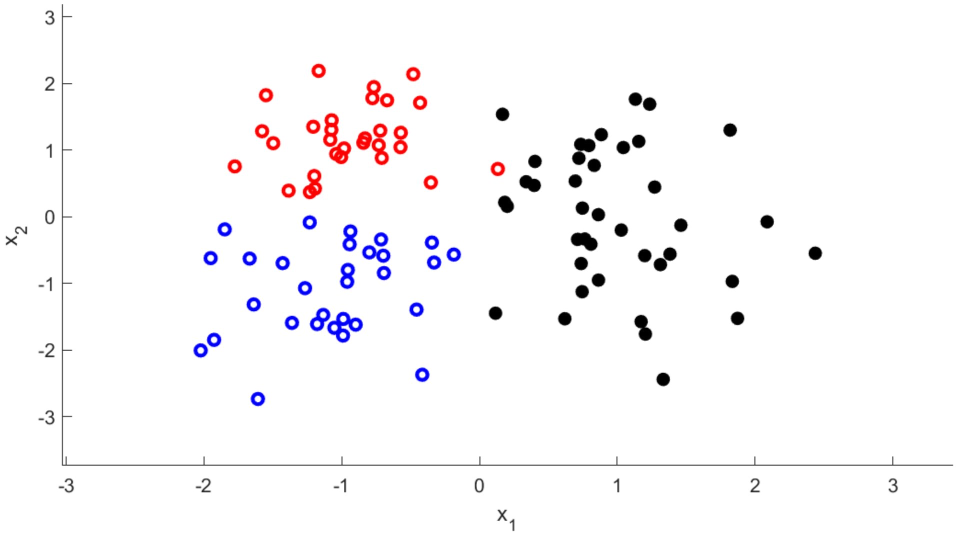
Assumptions in domain adaptation

Suppose you get a target data set without labels:



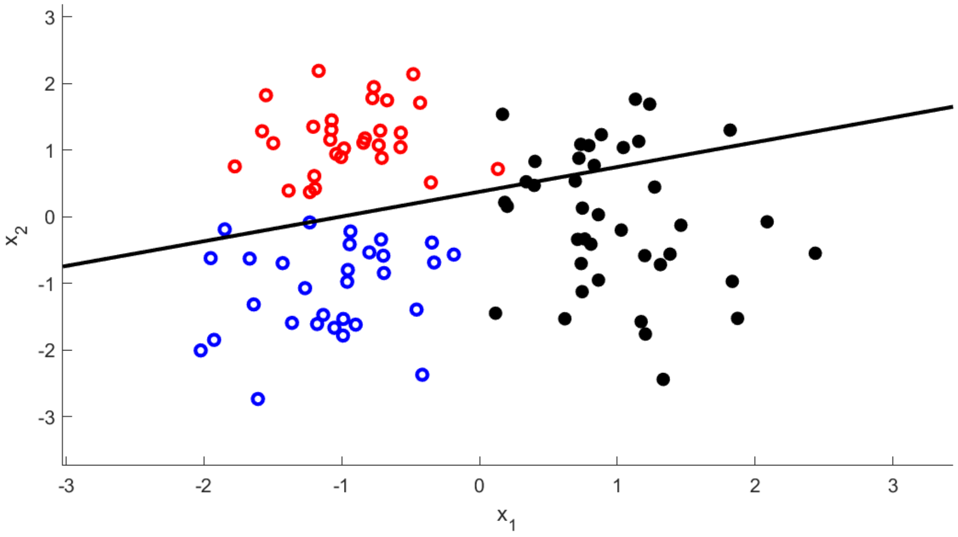
Assumptions in domain adaptation

You decide to use a source data set:



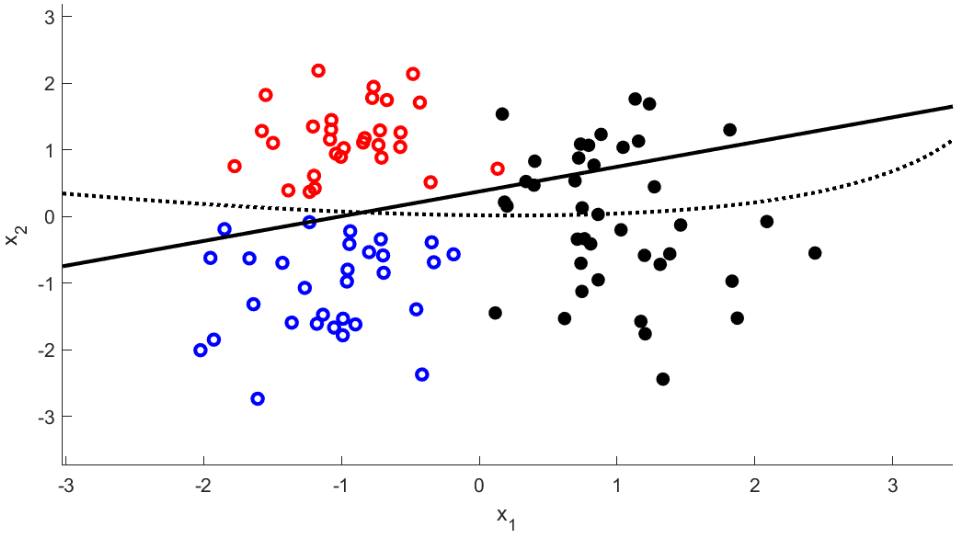
Assumptions in domain adaptation

You train a classifier on the source data and apply it to the target data:



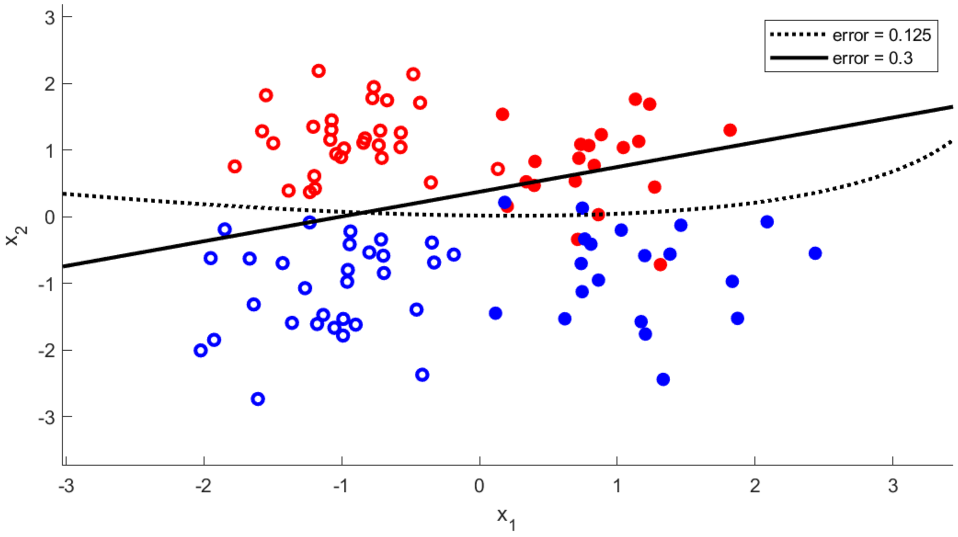
Assumptions in domain adaptation

You also decide to make an assumption on the relationship between domains and train a domain-adaptive classifier:



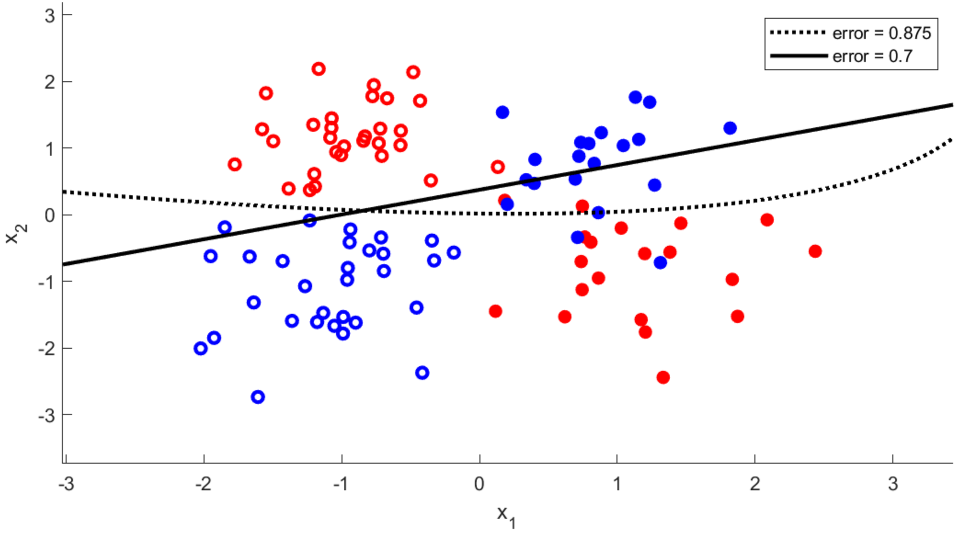
Assumptions in domain adaptation

If your assumption was correct, then your adaptive classifier is probably an improvement over the source-trained classifier.



Assumptions in domain adaptation

If your assumption was incorrect, then your adaptive classifier might perform worse:



Can we design an estimator that
will *always* improve over the source data estimator?

Target Robust Estimator

We propose an estimator $\hat{\theta}^{\mathcal{T}}$, for discriminant analyses, whose empirical risk $\hat{R}$ is *strictly less* than the risk of the source estimator $\hat{\theta}^{\mathcal{S}}$ on the given target data:

$$\hat{R}_{\mathrm{DA}}\big(\hat{\theta}^{\mathcal{T}} \mid z, u\big) \;<\; \hat{R}_{\mathrm{DA}}\big(\hat{\theta}^{\mathcal{S}} \mid z, u\big)$$

Target Robust Estimator

Suppose we have source data (x,y) and target data (z,u). The target labels u are unknown and must be predicted.

Target Robust Estimator

Suppose we have source data (*x,y*) and target data (*z,u*). The target labels *u* are unknown and must be predicted.

The source estimator is the discriminant analysis parameter estimator fitted to source data:

$$\hat{\theta}^{\mathcal{S}} = \underset{\theta \in \Theta}{\arg\min}\ \hat{R}_{\mathrm{DA}}(\theta \mid x, y)$$

Target Robust Estimator

Suppose we have source data (*x,y*) and target data (*z,u*). The target labels *u* are unknown and must be predicted.

The source estimator is the discriminant analysis parameter estimator fitted to source data:

$$\hat{\theta}^{\mathcal{S}} = \underset{\theta \in \Theta}{\arg\min} \ \hat{R}_{\mathrm{DA}}(\theta \mid x, y)$$

where the empirical risk is the negative log-likelihood of a Gaussian distribution for each class:

$$\hat{R}_{\mathrm{DA}}(\theta \mid x, y) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} -y_{ik} \log \left[\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k) \right]$$

Target Robust Estimator

If we had target labels *u*, we could measure the risk of the source estimator on the target data:

$$\hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, u)$$

Target Robust Estimator

If we had target labels *u*, we could measure the risk of the source estimator on the target data:

$$\hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, u)$$

With this, we could design an estimator that never performs worse than the source estimator:

Target Robust Estimator

If we had target labels *u*, we could measure the risk of the source estimator on the target data:

$$\hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, u)$$

With this, we could design an estimator that never performs worse than the source estimator:

$$\min_{\theta \in \Theta} \hat{R}_{\mathrm{DA}}(\theta \mid z, u) - \hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, u)$$

Target Robust Estimator

If we had target labels *u*, we could measure the risk of the source estimator on the target data:

$$\hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, u)$$

With this, we could design an estimator that never performs worse than the source estimator:

$$\min_{\theta \in \Theta} \hat{R}_{\mathrm{DA}}(\theta \mid z, u) - \hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, u)$$

This minimization procedure will either produce a θ with a lower risk or it will recover $\hat{\theta}^{\mathcal{S}}$. Values for θ that produce larger risks are not valid minimization solutions, so long as θ and $\hat{\theta}^{\mathcal{S}}$ are both drawn from the same parameter space Θ.

Target Robust Estimator

Since we don't have target labels *u*, we should prepare for the worst:

Target Robust Estimator

Since we don't have target labels *u,* we should prepare for the worst:

$$\hat{R}_{\mathrm{DA}}\left(\hat{\theta}^{\mathcal{S}} \mid z, u\right) \leq \max_{q}\ \hat{R}_{\mathrm{DA}}\left(\hat{\theta}^{\mathcal{S}} \mid z, q\right)$$

The labels q represent the labeling that produces the maximal risk for a given set of parameters.

Target Robust Estimator

Since we don't have target labels *u,* we should prepare for the worst:

$$\hat{R}_{\mathrm{DA}}\left(\hat{\theta}^{\mathcal{S}} \mid z, u\right) \leq \max_{q} \ \hat{R}_{\mathrm{DA}}\left(\hat{\theta}^{\mathcal{S}} \mid z, q\right)$$

The labels q represent the labeling that produces the maximal risk for a given set of parameters.

Applying this worst-case setting to the difference in risks, gives:

$$\hat{\theta}^{\mathcal{T}} = \arg\min_{\theta \in \Theta} \ \max_{q} \ \hat{R}_{\mathrm{DA}}(\theta \mid z, q) - \hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, q)$$

Target Robust Estimator

Since we don’t have target labels *u,* we should prepare for the worst:

$$\hat{R}_{\mathrm{DA}}\left(\hat{\theta}^{\mathcal{S}} \mid z, u\right) \leq \max_{q} \ \hat{R}_{\mathrm{DA}}\left(\hat{\theta}^{\mathcal{S}} \mid z, q\right)$$

The labels q represent the labeling that produces the maximal risk for a given set of parameters.

Applying this worst-case setting to the difference in risks, gives:

$$\hat{\theta}^{\mathcal{T}} = \underset{\theta \in \Theta}{\arg\min} \ \max_{q} \ \hat{R}_{\mathrm{DA}}(\theta \mid z, q) - \hat{R}_{\mathrm{DA}}(\hat{\theta}^{\mathcal{S}} \mid z, q)$$

We call this estimator the Target Robust (TR) estimator.

Target Robust Estimator

In the paper, we show that the Target Robust estimator will actually *always* produce a parameter estimate with a lower risk on the given target data.

Target Robust Estimator

In the paper, we show that the Target Robust estimator will actually *always* produce a parameter estimate with a lower risk on the given target data.

- This is because, in discriminant analyses, the parameter estimate is based on sample averages.

Target Robust Estimator

In the paper, we show that the Target Robust estimator will actually *always* produce a parameter estimate with a lower risk on the given target data.

- This is because, in discriminant analyses, the parameter estimate is based on sample averages.
- To produce *exactly* the same parameter estimates, $\hat{\theta}^{\mathcal{T}} = \hat{\theta}^{\mathcal{S}}$, the sample averages for the source and target data would have to be *exactly equal;*

$$\frac{1}{M} \sum_{j=1}^{M} z_j = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Target Robust Estimator

In the paper, we show that the Target Robust estimator will actually *always* produce a parameter estimate with a lower risk on the given target data.

- This is because, in discriminant analyses, the parameter estimate is based on sample averages.

- To produce *exactly* the same parameter estimates, $\hat{\theta}^{\mathcal{T}} = \hat{\theta}^{\mathcal{S}}$, the sample averages for the source and target data would have to be *exactly equal;*

$$\frac{1}{M} \sum_{j=1}^{M} z_j = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- The probability of drawing two sets of samples with exactly the same average is 0.

Discussion

- The Target Robust estimator does not depend on domain shift assumptions, such as "covariate shift" or "low-joint-domain-error".

Discussion

- The Target Robust estimator does not depend on domain shift assumptions, such as “covariate shift” or “low-joint-domain-error”.

- The results are only valid for *empirical risks,* not error rates.

Discussion

- The Target Robust estimator does not depend on domain shift assumptions, such as “covariate shift” or “low-joint-domain-error”.

- The results are only valid for *empirical risks,* not error rates.

- The results are only valid for the given target data, not future target samples.
    - Hence, the Target Robust estimator is *transductive* in nature.

Discussion

- The Target Robust estimator does not depend on domain shift assumptions, such as "covariate shift" or "low-joint-domain-error".

- The results are only valid for *empirical risks,* not error rates.

- The results are only valid for the given target data, not future target samples.
    - Hence, the Target Robust estimator is *transductive* in nature.

- If the source estimator performs below chance, then there is no guarantee that the Target Robust estimator will perform above chance level.
    - At least, not without additional asssumptions.

Wouter M. Kouw
Signal Processing Systems group
TU Eindhoven

Marco Loog
Pattern Recognition lab
TU Delft

Thank you for your time.

https://arxiv.org/abs/1806.09463
https://github.com/wmkouw/tcpr/
https://wmkouw.github.io/