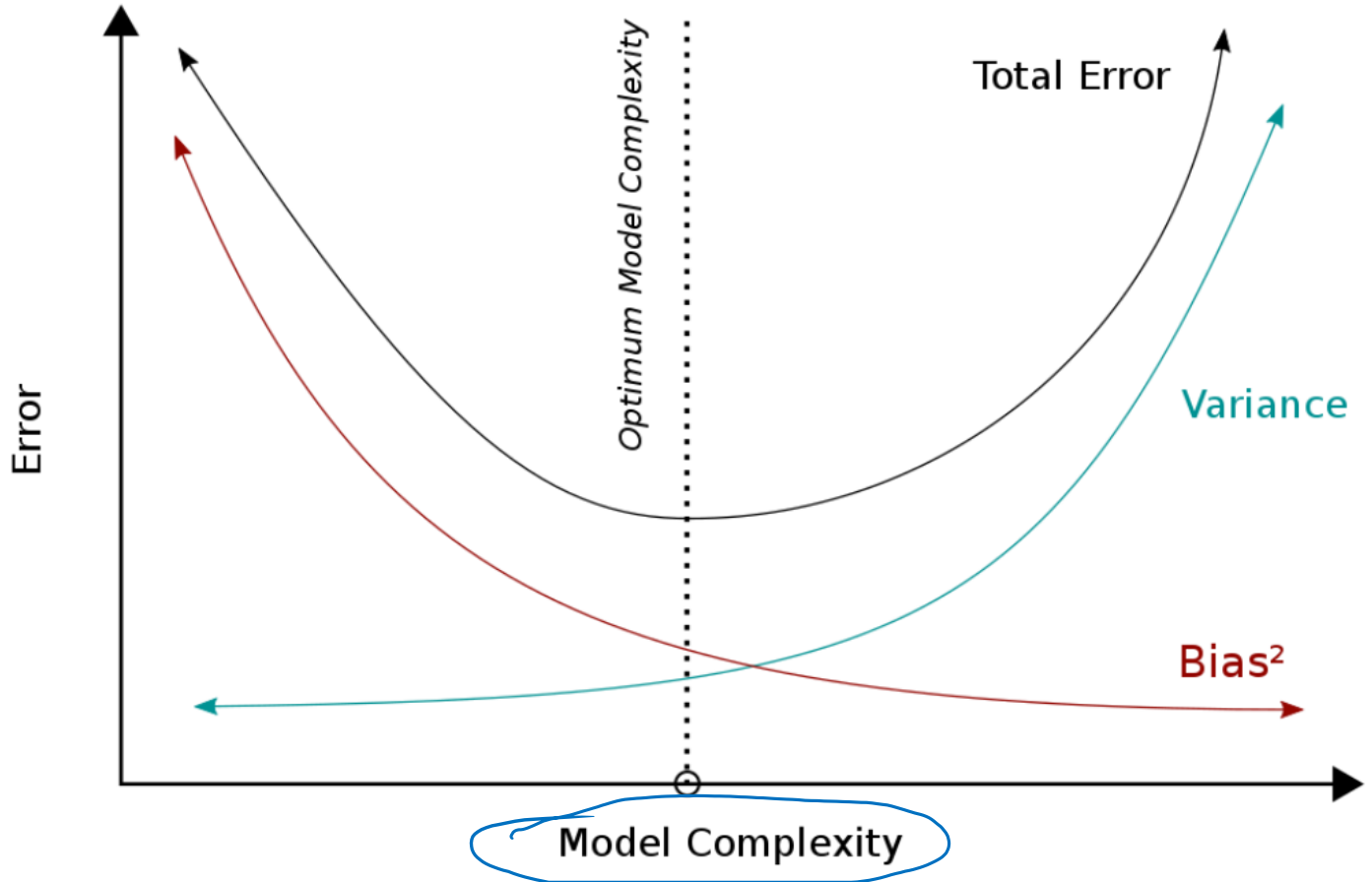
Generalization

Mostafa Tavassolipour

Bias-Variance Tradeoff

- Dimensionality reduction decreases the model complexity.
- Adding feature increases the model complexity.
- The **more complex** model, the **larger sample size** is needed.



Error = Bias^2 + Variance + Bayes Error

VC dimension

- VC: Vapnik-Chervonenkis

Statistical Learning Theory

with prob. $1-\eta$

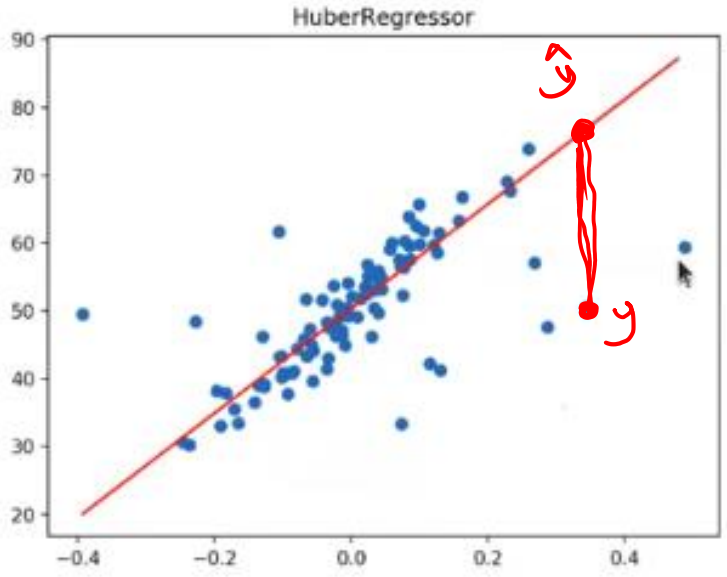
$$\text{test error} \leqslant \text{training error} + \sqrt{\frac{1}{N}\left[D\left(\log\left(\frac{2N}{D}\right)+1\right)-\log\left(\frac{\eta}{4}\right)\right]}$$

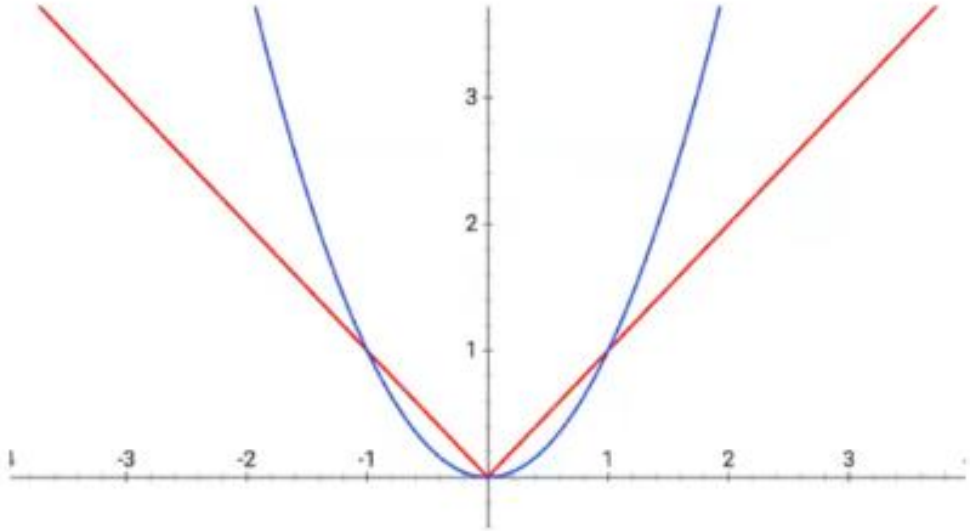
- N: number of training samples
- D: VC dimension
- η: a probability value near zero

How to impose inductive biases?

- Model architecture
- Input features (Feature Extraction)
- Loss function
- Data Augmentation
- Optimizer

Loss Function

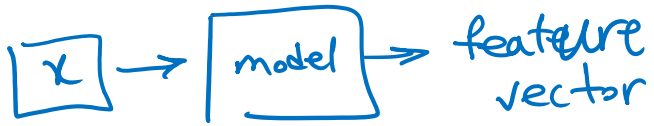




$\ell_1 : \quad |(y-\hat{y})|$

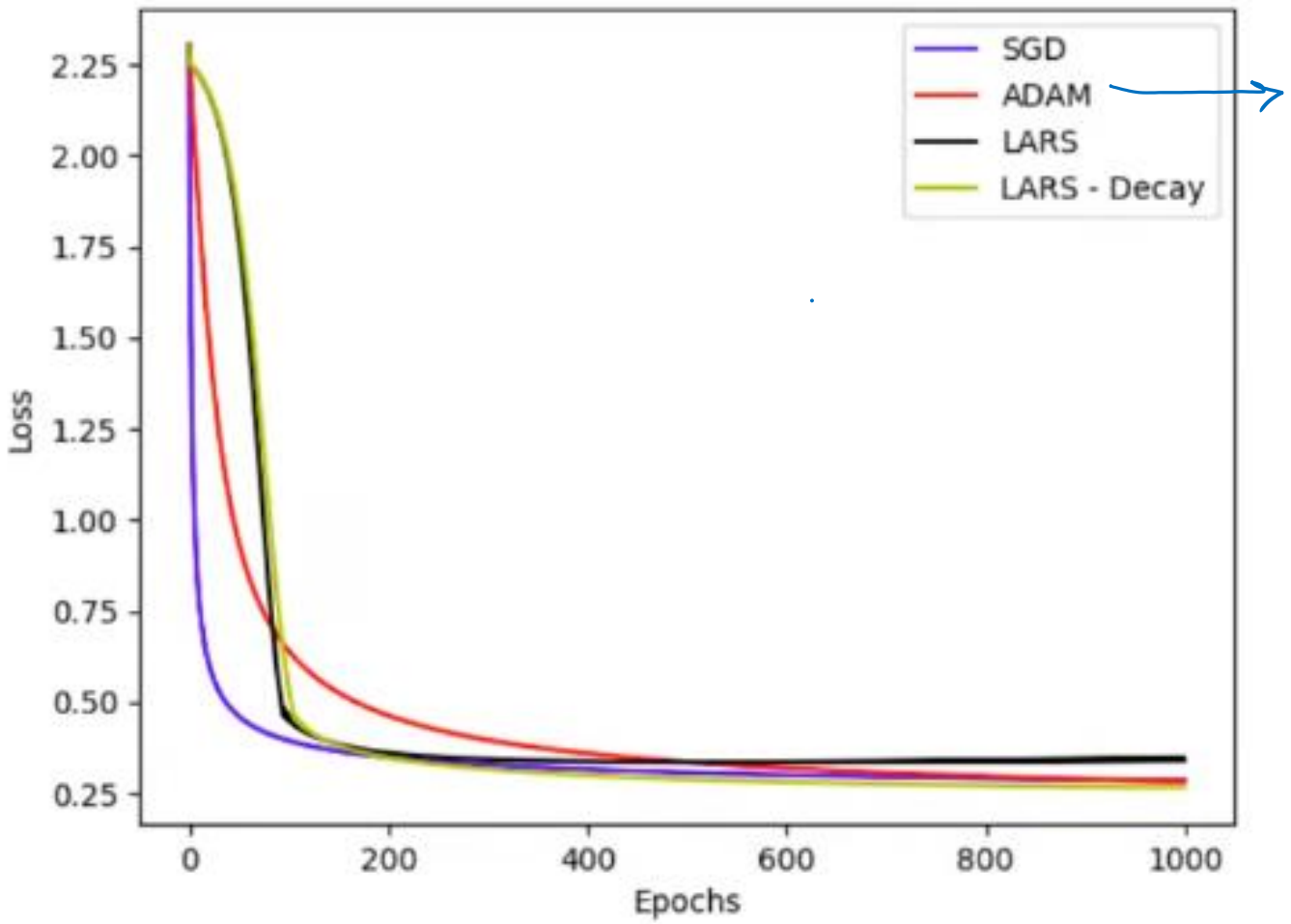
$\ell_2 : \quad (y-\hat{y})^2$

Data Augmentation



Optimizer

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L$$



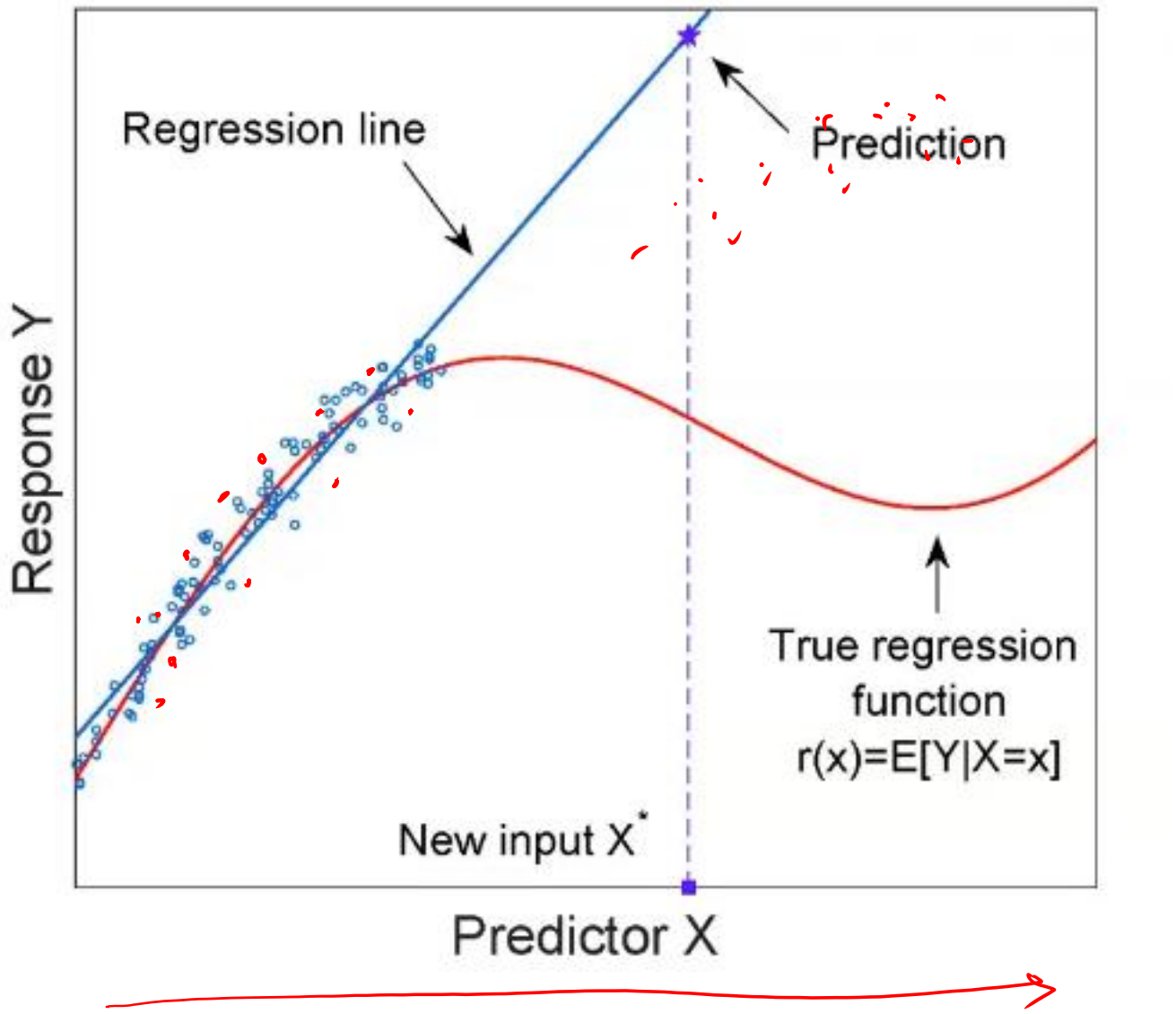
Generalization vs Domain Generalization

Domain Adaptation

- Generalization:
  - Refers to the ability of a machine learning model to accurately predict outcomes for new, **unseen** data points that come from the **same distribution** as the training data.

- Domain Generalization:
  - domain generalization is an extension of generalization that focuses on the ability of a model to generalize across **different domains** or datasets.
  - The **goal** is to train a model that can perform well not only on the training dataset but also on other datasets that have **different data distributions** but come from the same overall domain.

Generalization



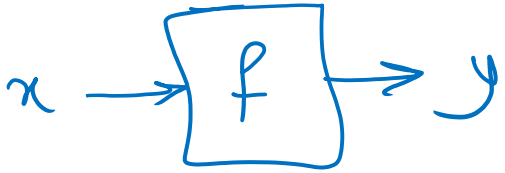
Transfer Learning vs. Domain Adaptation vs. Domain Generalization

- **Transfer learning** involves leveraging knowledge gained from a source task or domain to improve learning and performance in a related target task or domain.

- **Domain adaptation** aims to adapt a model trained on a source domain to perform well on a different target domain. It deals with scenarios where the source and target domains have different distributions or characteristics.

Task and Domain

$p(x, y) = p(x)\, p(y|x)$



$x \in X$, $p(x)$ } domain

$y \in Y$, $p(y|x)$ } task

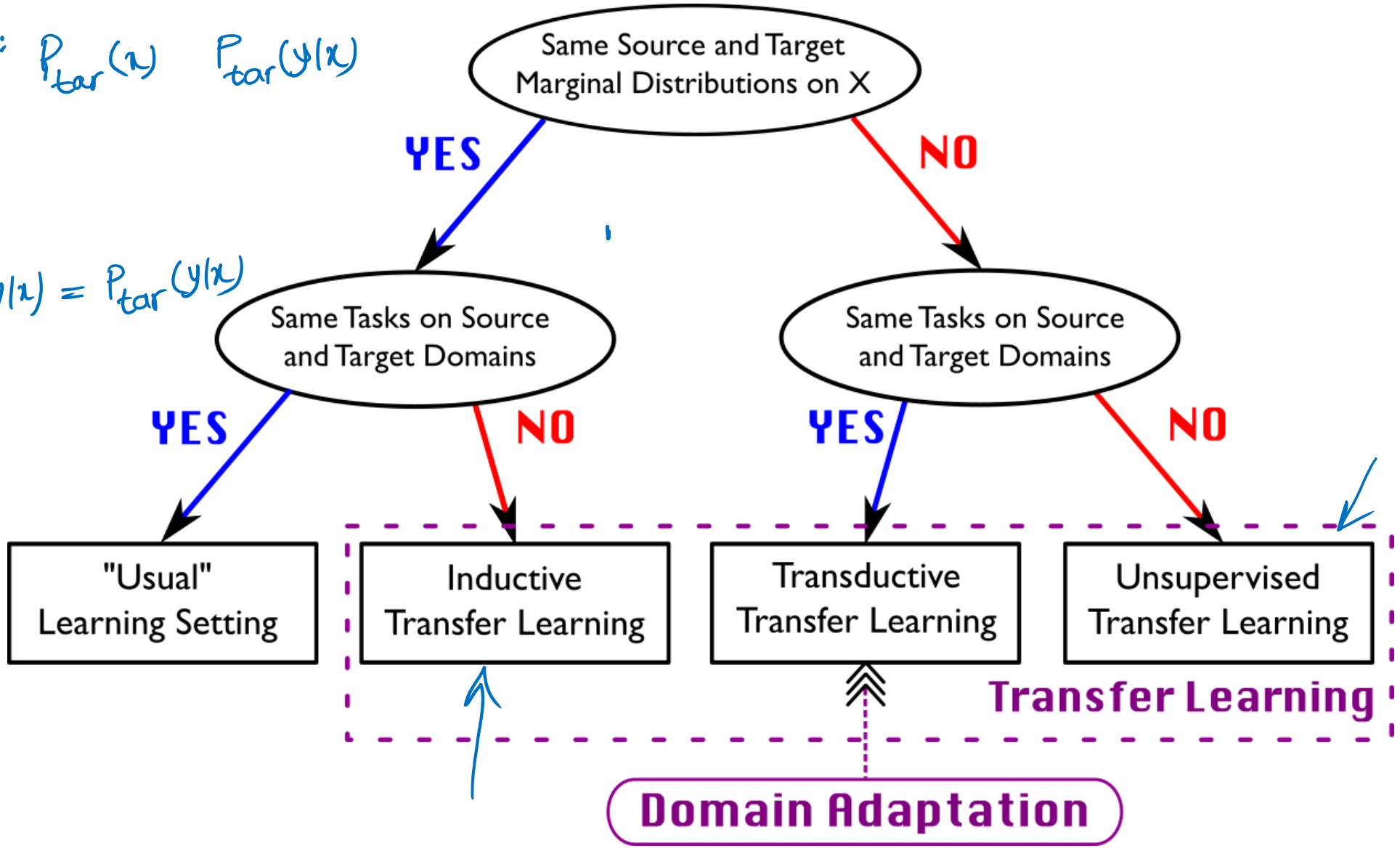
Transfer Learning vs. Domain Adaptation vs. Domain Generalization

source: $P_{src}(x)$ $P_{src}(y|x)$

target: $P_{tar}(x)$ $P_{tar}(y|x)$

$P_{src}(x) = P_{tar}(x)$

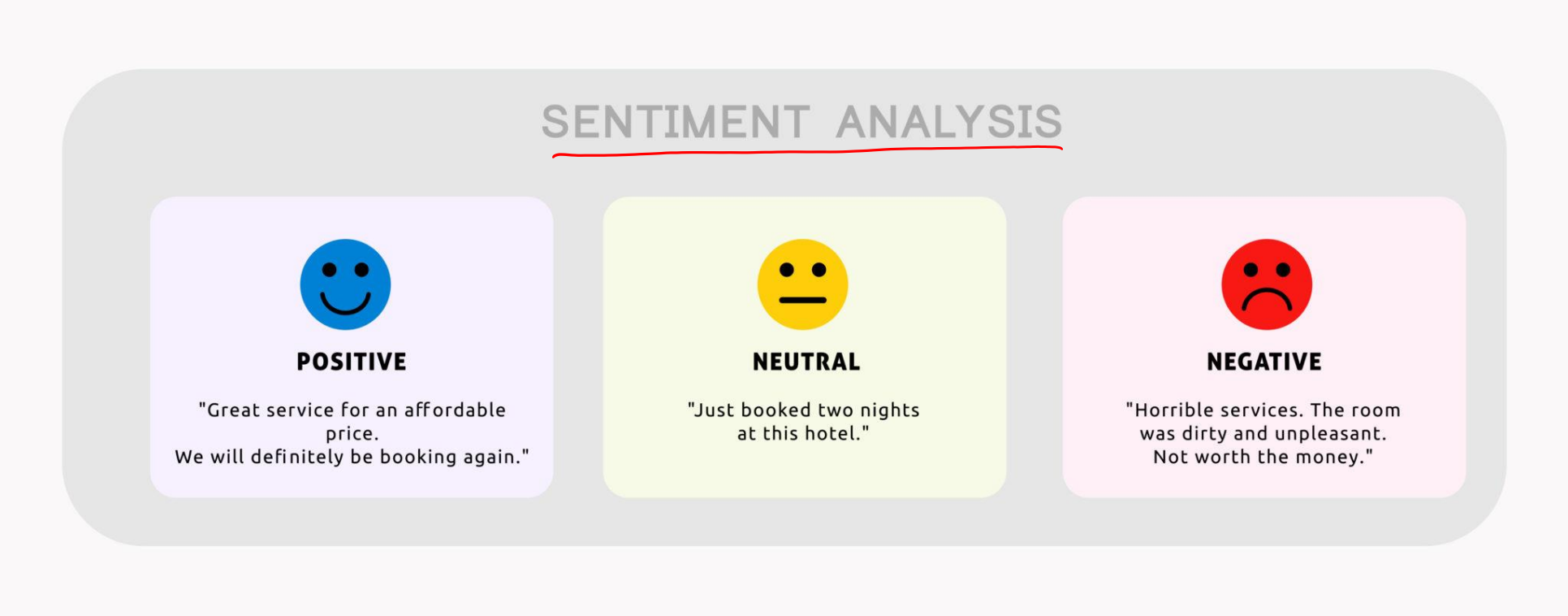
$P_{src}(y|x) = P_{tar}(y|x)$



Inductive Transfer Learning: Example

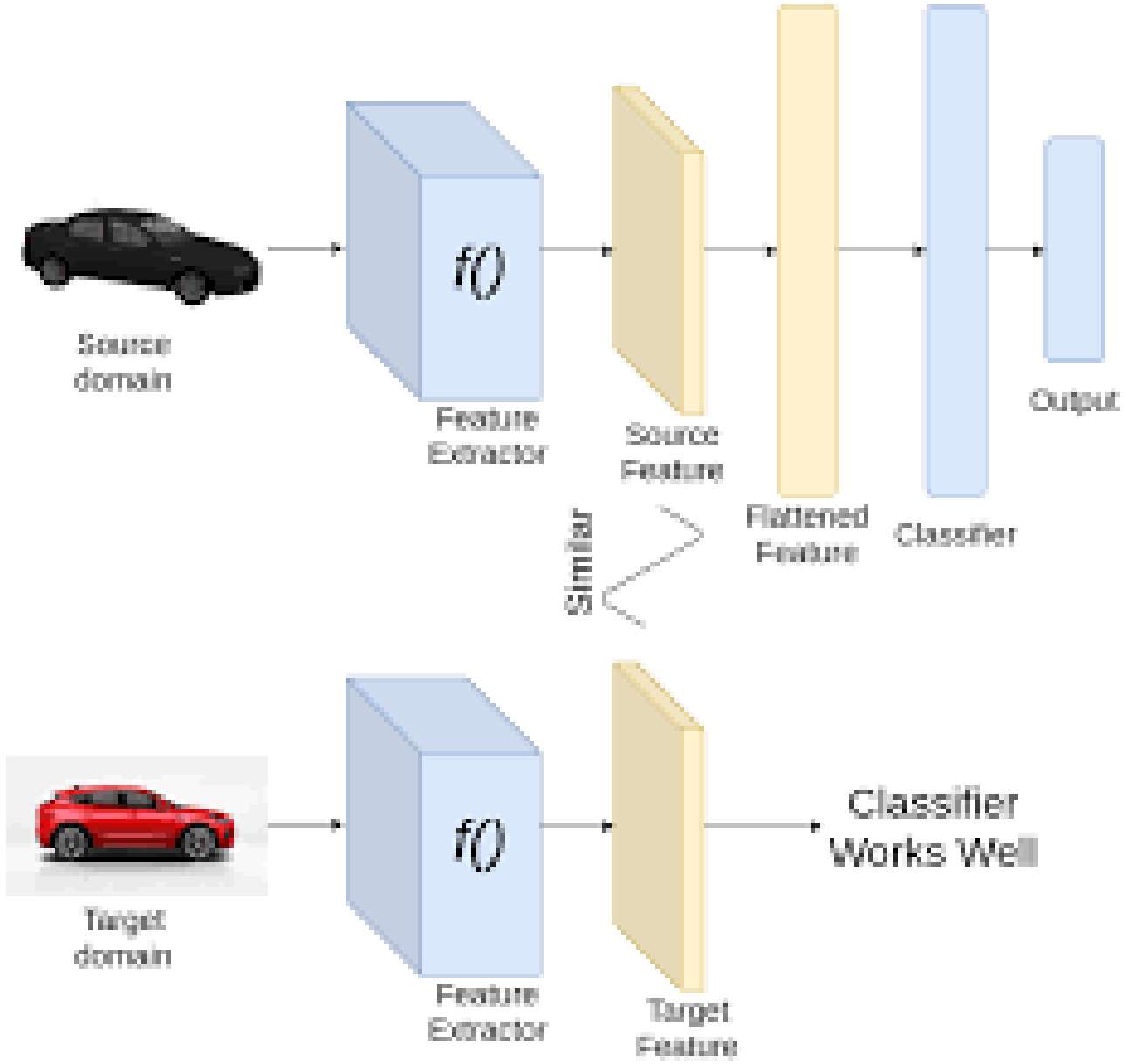
- Same domain
- Different tasks

BERT



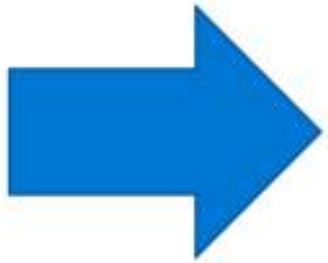
Domain Adaptation: Example

- Different Domains
- Same task



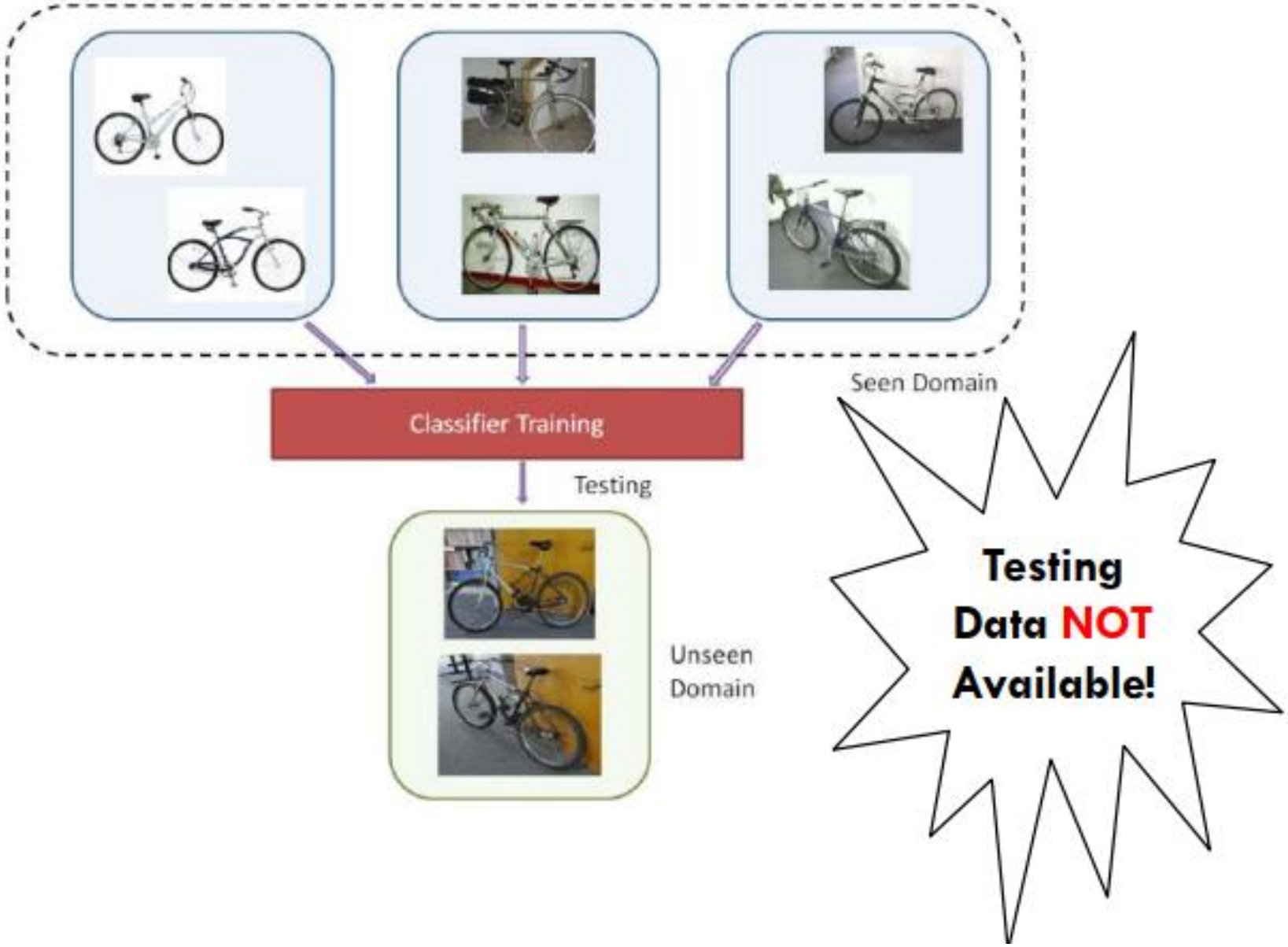
Domain Adaptation: Example 2

ImageNet



CiFAR100

Domain Generalization

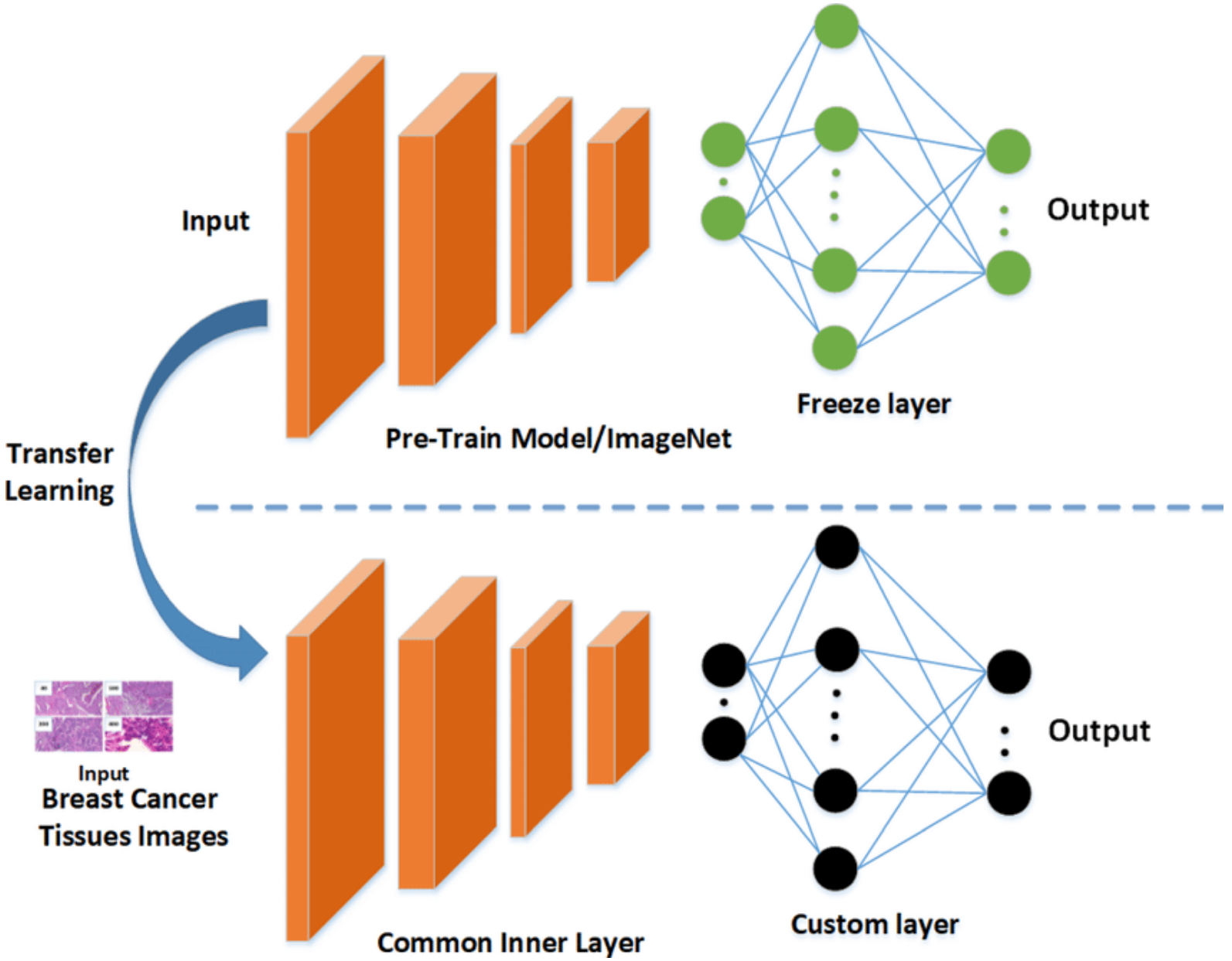


Unsupervised Transfer Learning: Example

- Different domains
- Different tasks

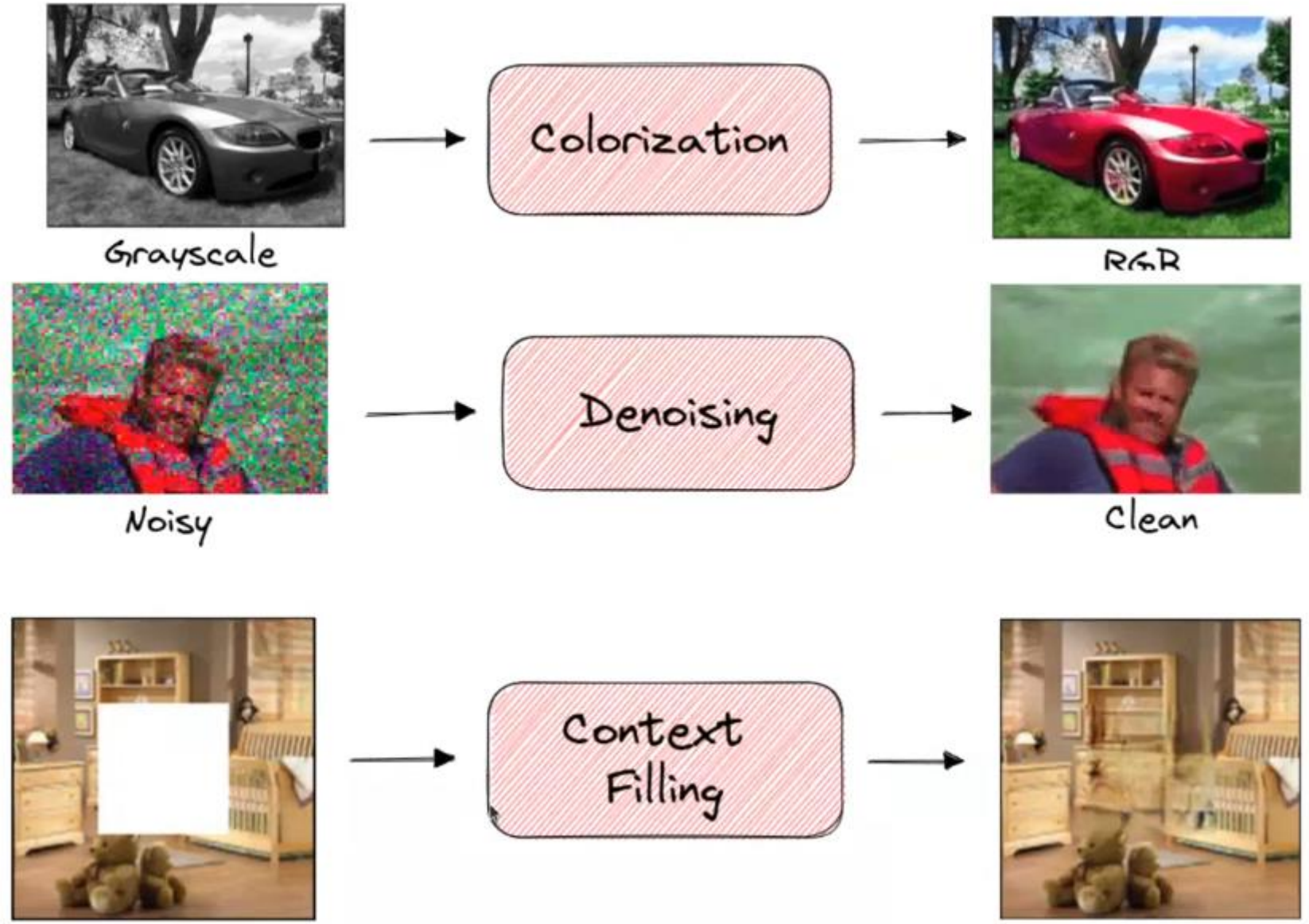
$x \in X$

$P(y|x)$

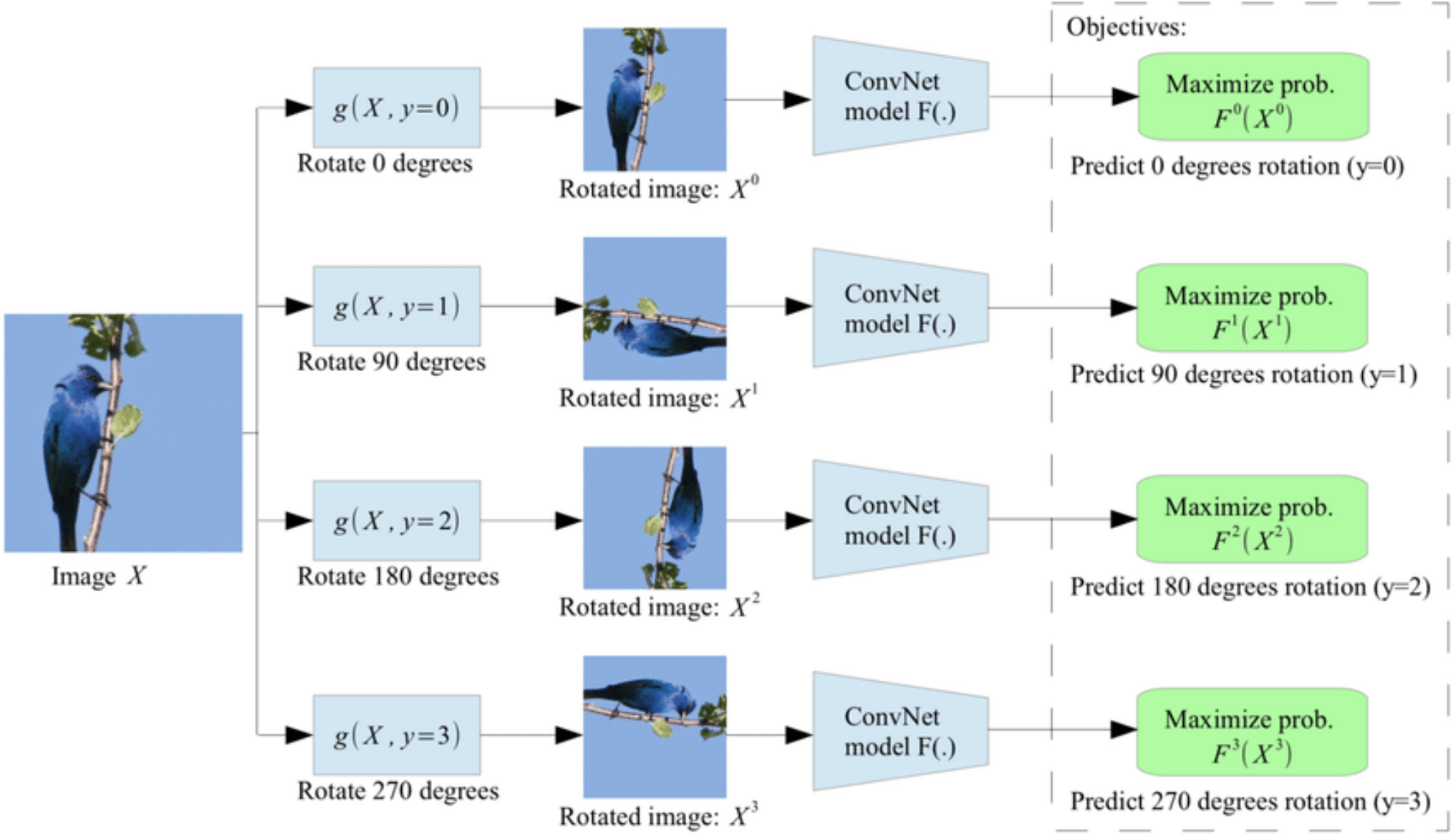


Self-Supervision

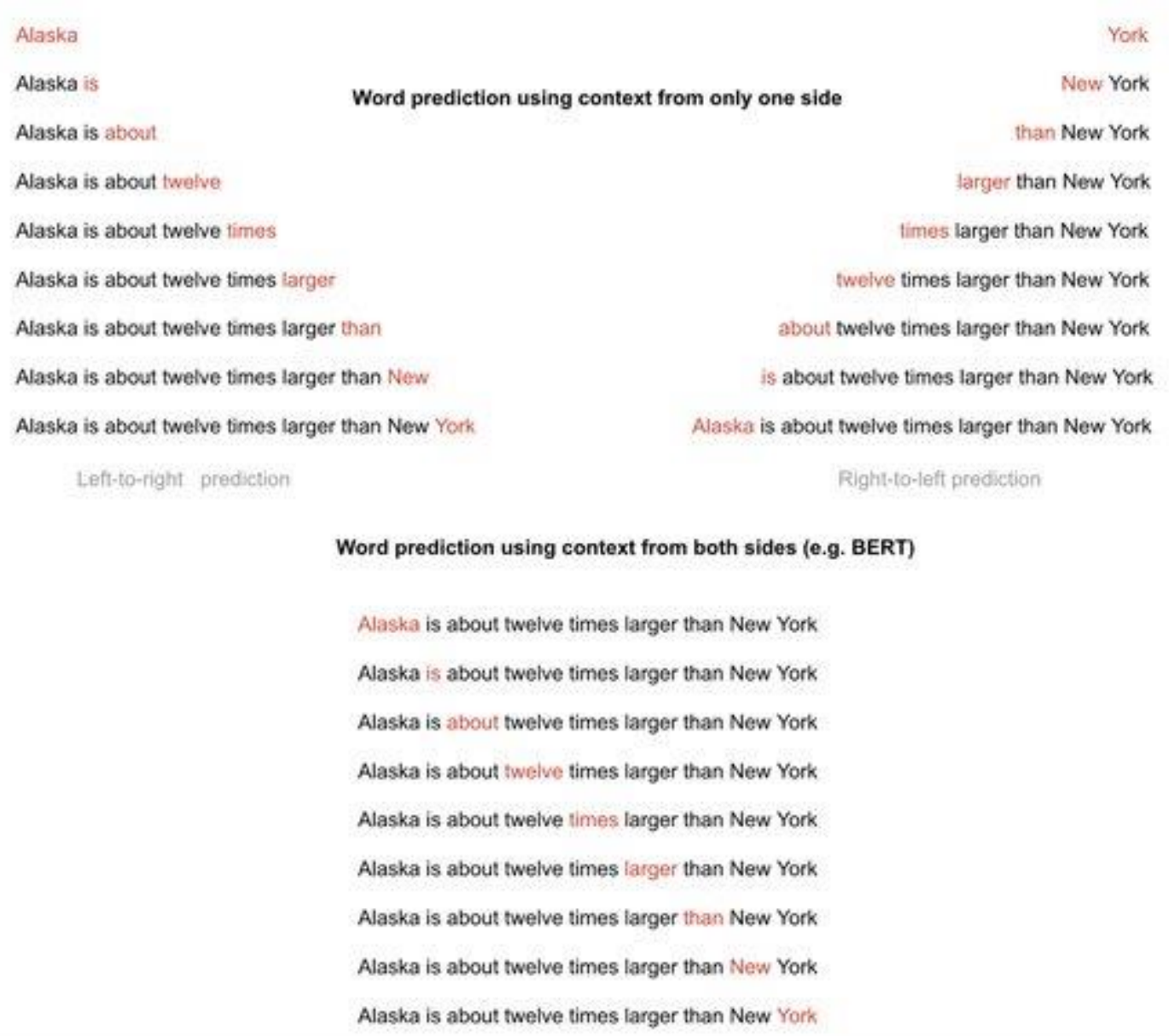
pre-text task



Self-Supervision for Image classification



Masked Language Models



Self-Supervision and Transfer Learning: Sentiment Analysis