Transfer Learning

Transfer Learning - Overview

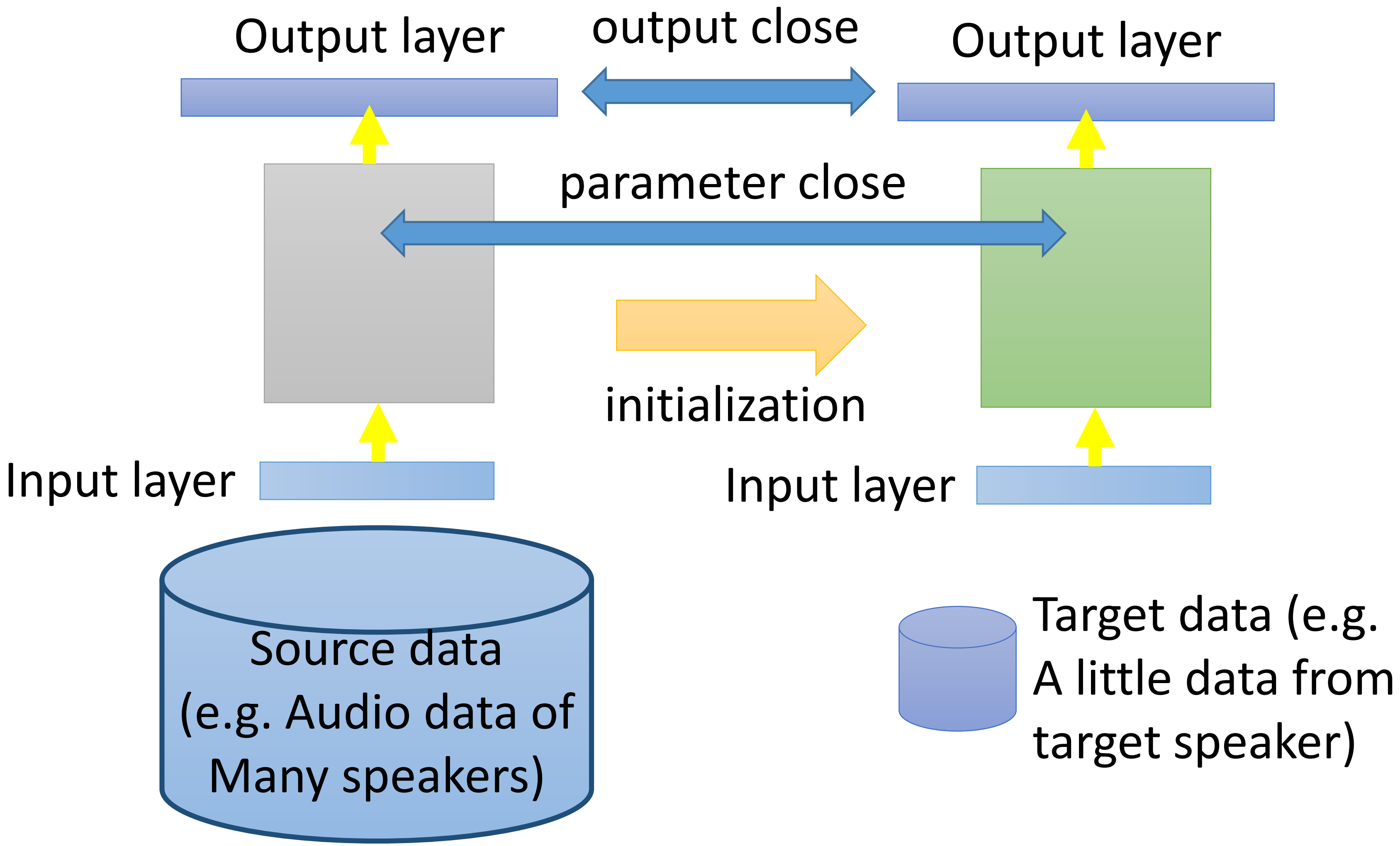
		Source Data (not directly related to the task)	
		labelled	unlabeled
Target Data	labelled	Fine-tuning Multitask Learning	
	unlabeled		

Model Fine-tuning

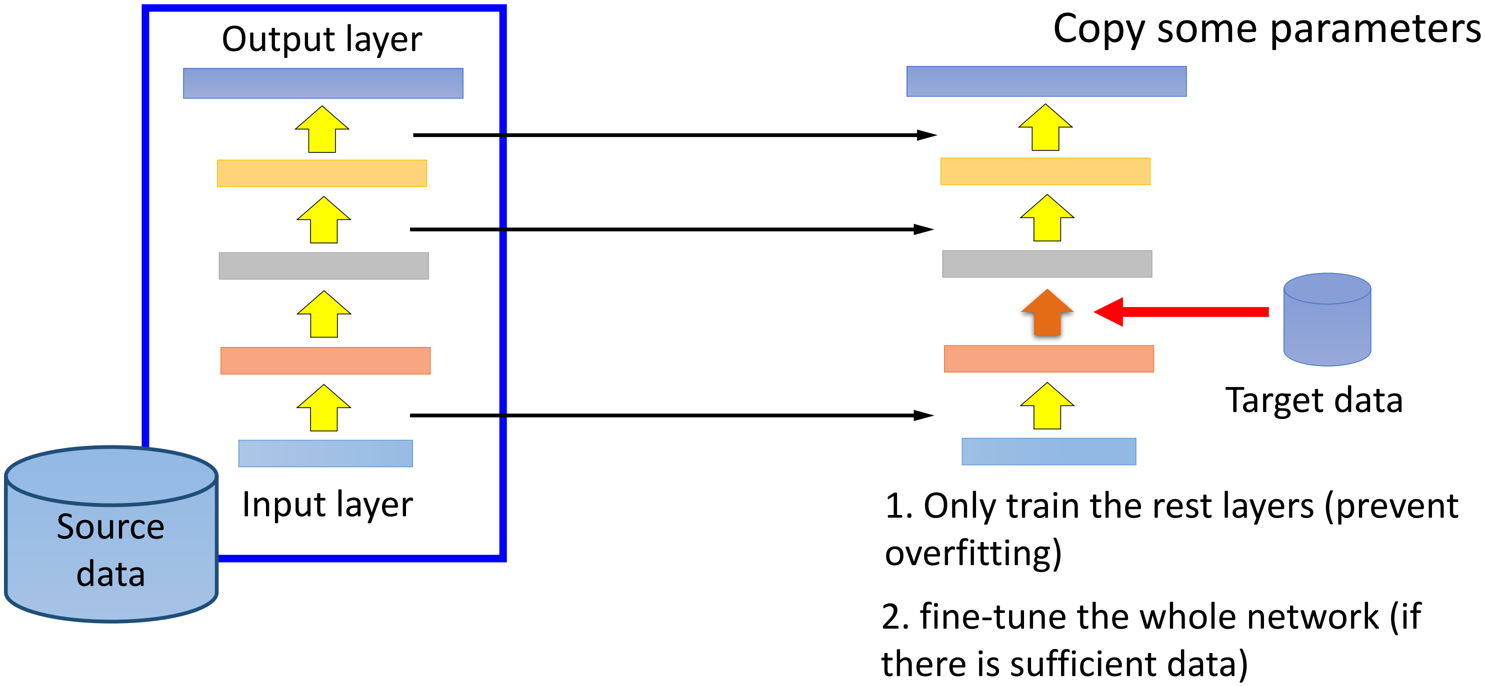
One-shot learning: only a few examples in target domain

- Task description
  - Source data: (x^s, y^s) ⇦ A large amount
  - Target data: (x^t, y^t) ⇦ Very little
- Example: (supervised) speaker adaption
  - Source data: audio data and transcriptions from many speakers
  - Target data: audio data and its transcriptions of specific user
- Idea: training a model by source data, then fine-tune the model by target data
  - Challenge: only limited target data, so be careful about overfitting

Conservative Training



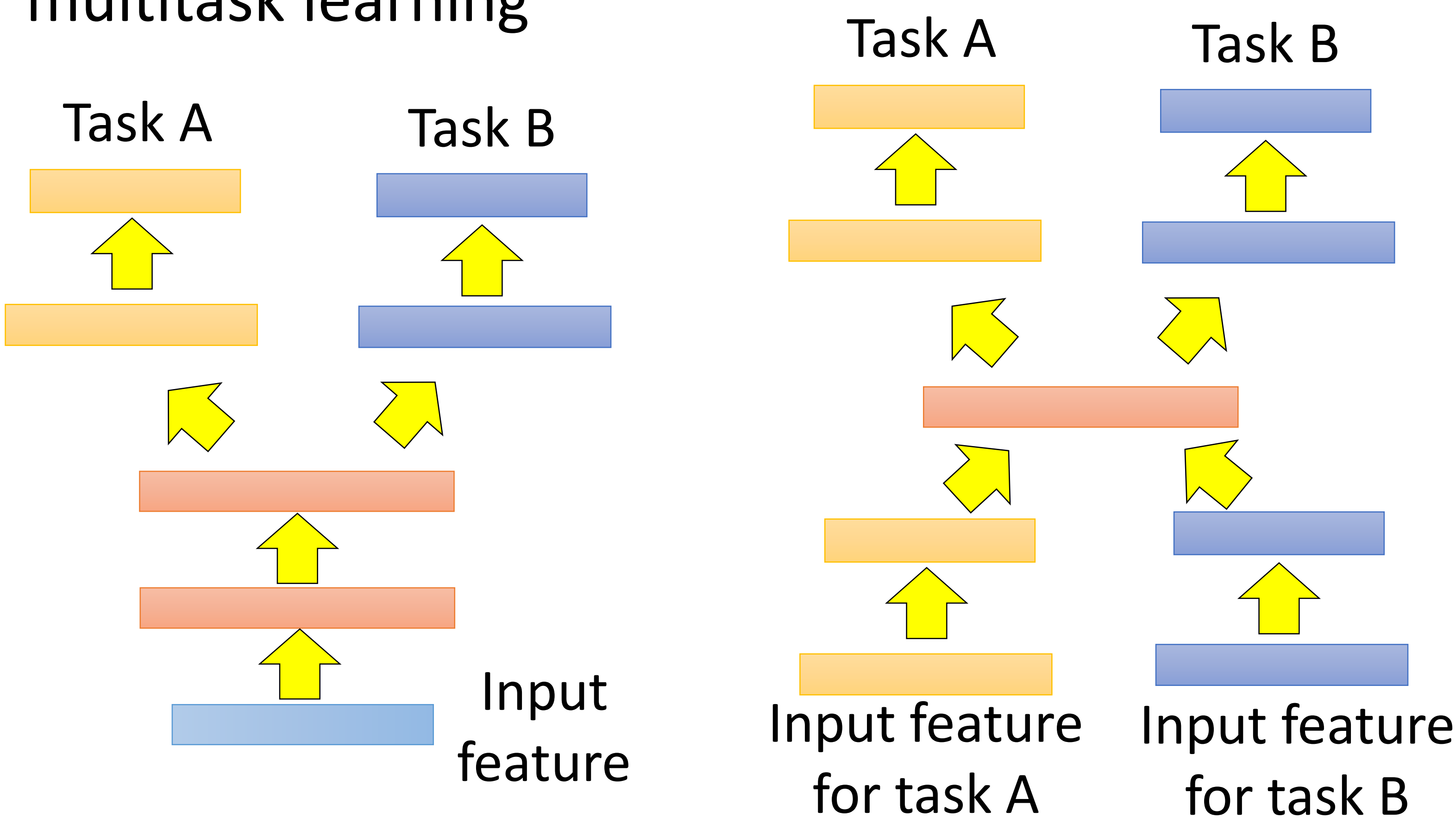
Layer Transfer



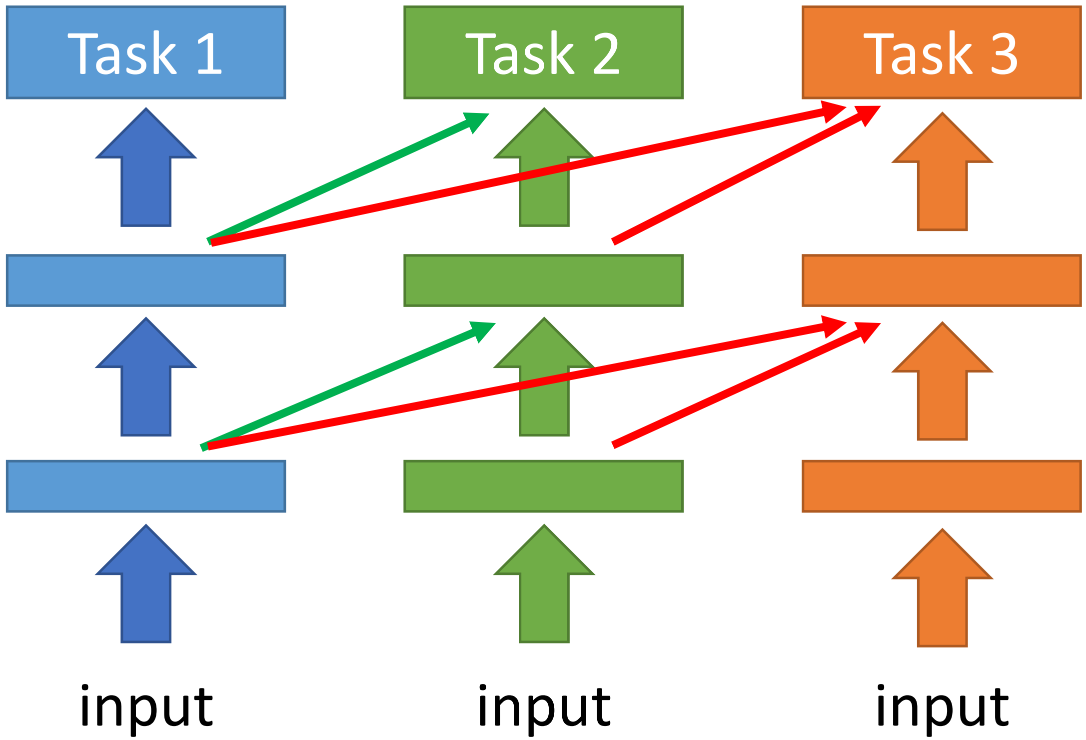
- Which layer can be transferred (copied)?
  - Speech: usually copy the last few layers
  - Image: usually copy the first few layers

Multitask Learning

- The multi-layer structure makes NN suitable for multitask learning



Progressive Neural Networks



Andrei A. Rusu, Neil C. Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, Raia Hadsell, “Progressive Neural Networks”, arXiv preprint 2016

Transfer Learning - Overview

		Source Data (not directly related to the task)	
		labelled	unlabeled
Target Data	labelled	Fine-tuning Multitask Learning	
	unlabeled	Domain-adversarial training Zero-shot learning	

Task description

- Source data: (x^s, y^s) → Training data
- Target data: (x^t) → Testing data

Same task, mismatch

MNIST

SOURCE

with label

TARGET

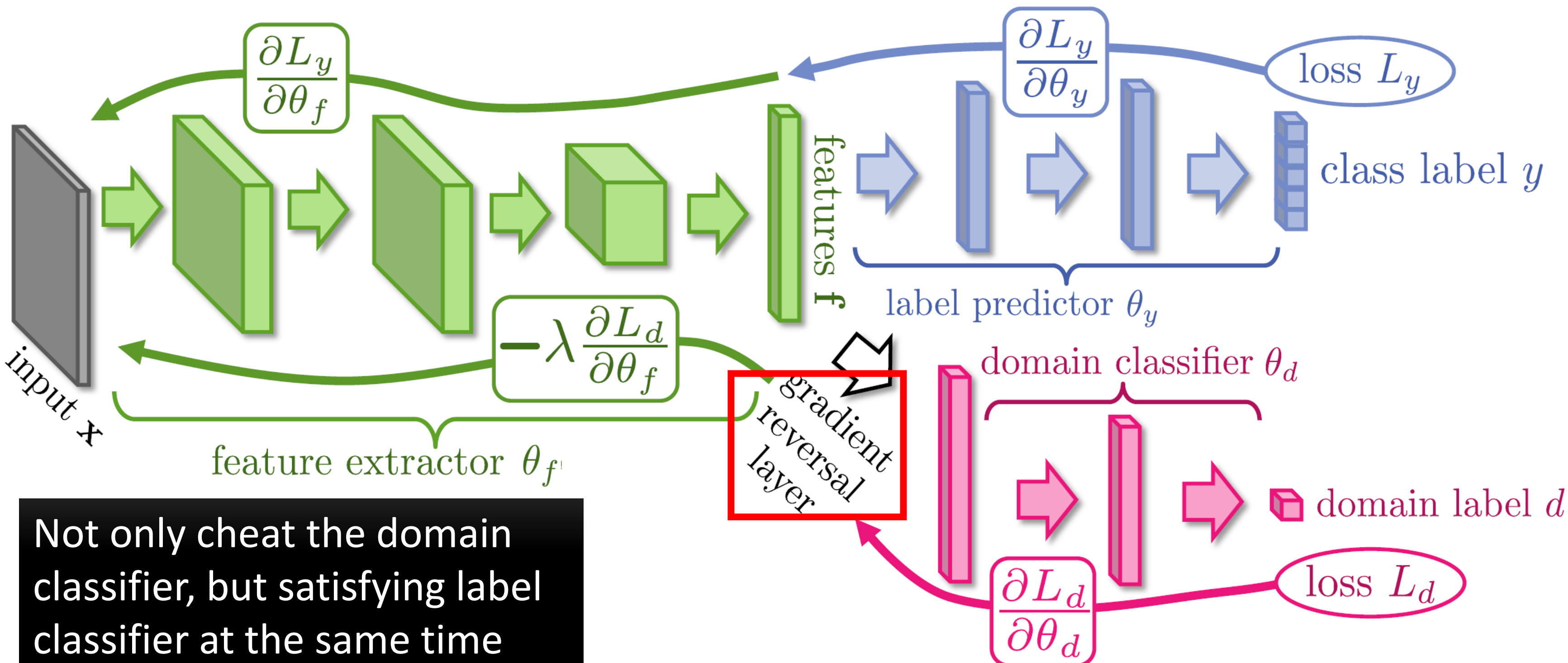
without label

MNIST-M

Domain-adversarial training

Maximize label classification accuracy + minimize domain classification accuracy

Maximize label classification accuracy



Not only cheat the domain classifier, but satisfying label classifier at the same time

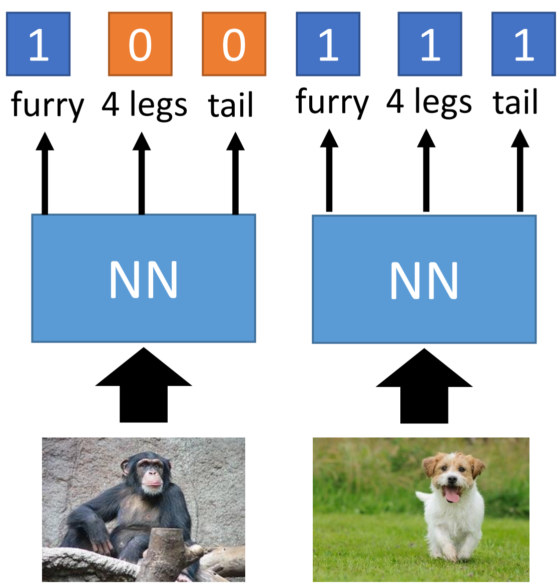
Maximize domain classification accuracy

Zero-shot Learning

- Source data: (x^s, y^s) → Training data
- Target data: (x^t) → Testing data
- Representing each class by its attributes

Different tasks

Training



Database

attributes

class

	furry	4 legs	tail	…
Dog	O	O	O	
Fish	X	X	O	
Chimp	O	X	X	
…				

sufficient attributes for one to one mapping

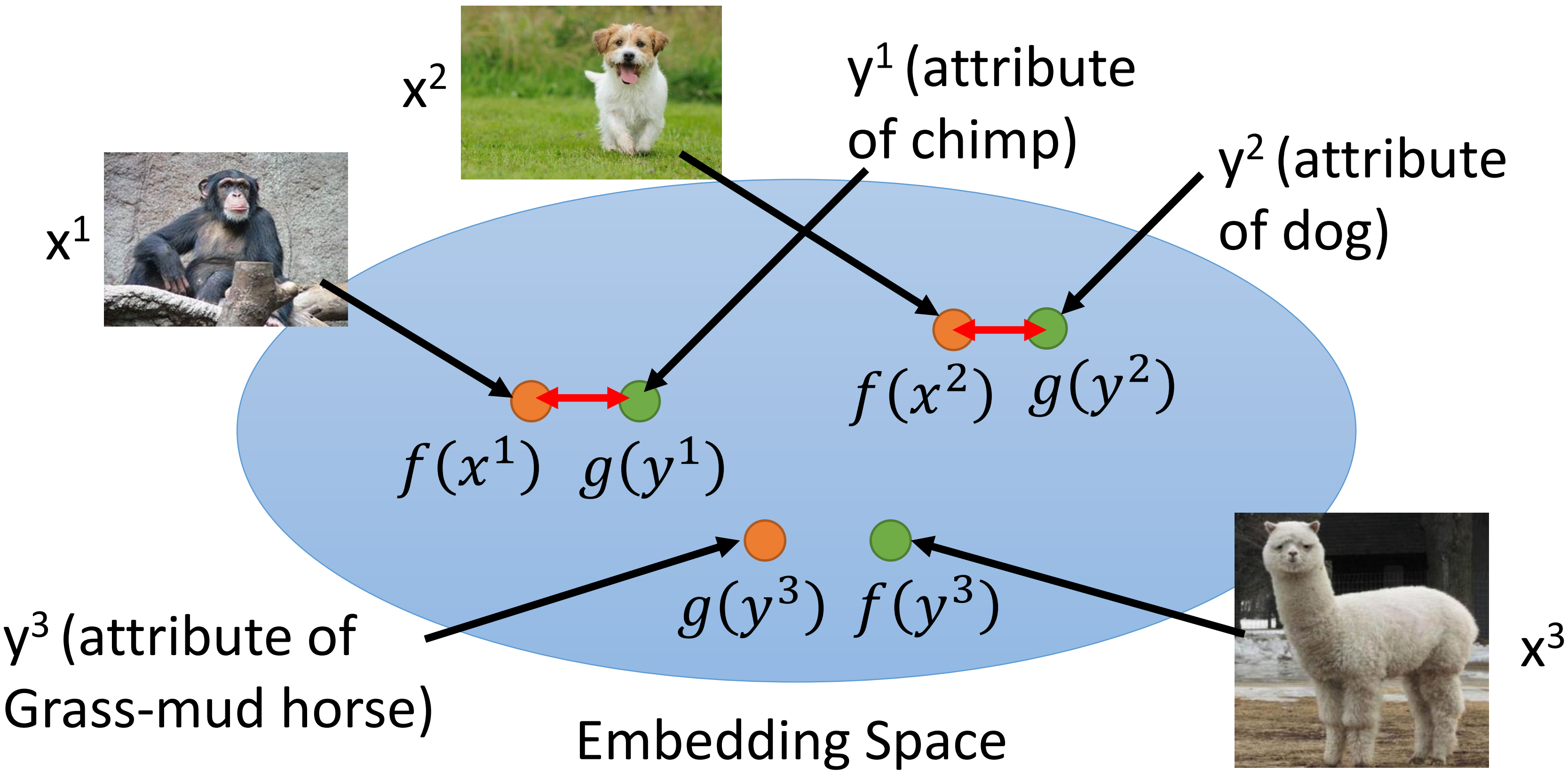
Zero-shot Learning

$f(*)$ and $g(*)$ can be NN.

Training target:

$f(x^n)$ and $g(y^n)$ as close as possible

- Attribute embedding



Zero-shot Learning

$$f^*, g^* = arg \min_{f,g} \sum_n \|f(x^n) - g(y^n)\|_2$$ Problem?

$$f^*, g^* = arg \min_{f,g} \sum_n max\Big(0, k - f(x^n) \cdot g(y^n) + \max f \;\; x^m \;\; \cdot g$$

$$, g^* = arg \min_{f,g} \sum_n max \;\; 0, m - f \;\; x^n \;\; \cdot g \;\; x^n \;\; + \max_{m \neq n} f(x^n) \cdot g(y^m)\Big)$$

Margin you defined

Zero loss: $k - f(x^n) \cdot g(y^n) + \max_{m \neq n} f(x^n) \cdot g(y^m) < 0$

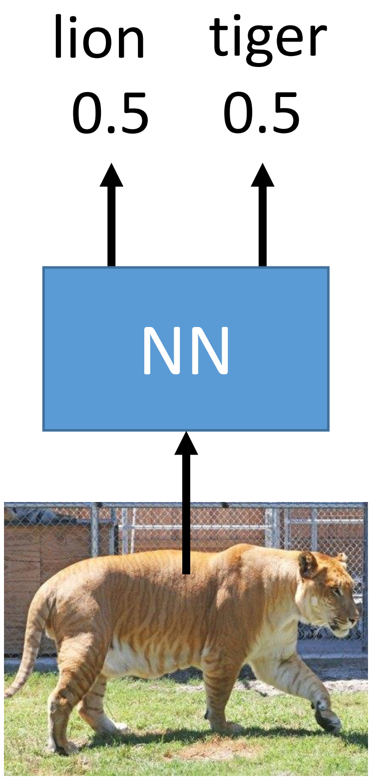
$$\underline{f(x^n) \cdot g(y^n)} - \underline{\max_{m \neq n} f(x^n) \cdot g(y^m)} > k$$

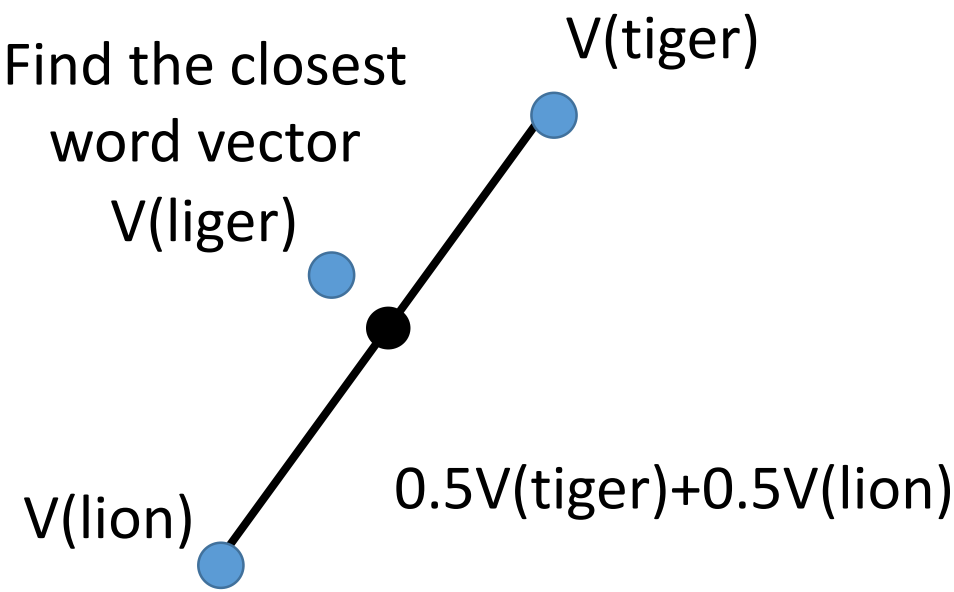
$f(x^n)$ and $g(y^n)$ as close

$f(x^n)$ and $g(y^m)$ not as close

Zero-shot Learning

- Convex Combination of Semantic Embedding





Only need off-the-shelf NN for ImageNet and word vector

Transfer Learning - Overview

		Source Data (not directly related to the task)	
		labelled	unlabeled
Target Data	labelled	Fine-tuning Multitask Learning	Self-taught learning Rajat Raina , Alexis Battle , Honglak Lee , Benjamin Packer , Andrew Y. Ng, Self-taught learning: transfer learning from unlabeled data, ICML, 2007
	unlabeled	Domain-adversarial training Zero-shot learning	Self-taught Clustering Wenyuan Dai, Qiang Yang, Gui-Rong Xue, Yong Yu, "Self-taught clustering", ICML 2008

Different from semi-supervised learning