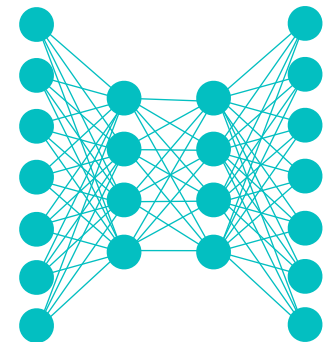


# Lecture Notes for **Neural Networks and Machine Learning**



Transfer Learning  
Deep Transfer Learning



# Logistics and Agenda

- Logistics
  - Today is mostly review, will go quickly!
  - Change Lab One?
  - Updated Student Presentation List
- Agenda
  - ConceptNet (if needed)
  - Transfer Learning Overview
  - Transfer Learning in Deep Learning
  - Demo
- Next Time:
  - Transformers for Text and Vision
- Next Next Time
  - Self-Supervised Learning and Consistency Loss





# How to Make a Racist AI without Really Trying



Robyn Speer, 2017

<http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/>

## Debiasing: Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Bolukbasi et al., NeurIPs 2016

<https://arxiv.org/pdf/1607.06520.pdf>

## ConceptNet 5.5: An Open Multilingual Graph of General Knowledge

Speer et al., AAAI 2017

<https://arxiv.org/pdf/1612.03975.pdf>



Rachael Tatman @rctatman · 18h

I first got interested in ethics in NLP/ML because I was asking "does this system work well for everyone". It's a good question, but there's a more important one:

Who is being harmed and who is benefiting from this system existing in the first place?




# Transfer Learning Overview

Transfer learning be like



# Transfer Learning

- Transfer knowledge from a source prediction task to a target prediction task
  - without any regard for performing well on source task
- **Original:** Neural Information Processing 1995 (NeuRiPs)
  - Workshop on “Learning to Learn”
    - ◆ How to effectively retain and reuse previously learned knowledge
  - Originally used in Markov chain and Bayesian networks (keeping n-grams, *etc.*)
- **Key idea:** Humans can generalize what they learn to almost any domain, can we mimic this behavior with ML?
  -  Ugh, really ML people?



## Ian Goodfellow's Definition:

*“Transfer learning refers to any situation where what has been learned in one setting is exploited to improve generalization in another setting.”*



# Transfer Learning: Large Umbrella

- Appears under many associations in the literature:
  - Learning to learn / Life-long learning
  - Knowledge transfer / Inductive transfer
  - Multi-task learning
  - Knowledge consolidation
  - Context-sensitive learning
  - Knowledge-based inductive bias
  - Meta learning
  - Incremental learning
  - Cumulative learning
  - Domain adaptation



# Precise Definition of Transfer Learning

$$X = x_1, x_2, \dots, x_N \in \mathcal{X}$$

$$Y = y_1, y_2, \dots, y_N \in \mathcal{Y}$$

$$\mathcal{D} = \{\mathcal{X}, p(X)\}$$

Domain      Feature Space      Probability Observation

$$\mathcal{T} = \{\mathcal{Y}, p(Y|X)\}$$

Task      Label Space      Learned Probability

- $\mathcal{D}$  Domain defines the features used and probability
- $\mathcal{X}$  is the space of all possible features
- $p(X)$  is probability of observing specific instances in  $\mathcal{X}$ 
  - Typically **intractable** to calculate (generative)
- $\mathcal{T}$  Task is within a domain, defining labels and model
- $\mathcal{Y}$  is space of all possible labels
- $p(Y|X)$  probability of observing specific label given the specific feature:
  - **Not intractable** (discriminative)





# Definition with Examples

$$X = x_1, x_2, \dots, x_N \in \mathcal{X}$$

$$Y = y_1, y_2, \dots, y_N \in \mathcal{Y}$$

$$\mathcal{D} = \{\mathcal{X}, p(X)\}$$

Domain	Feature Space	Probability Observation
--------	---------------	-------------------------

- Image Pixels
- Sensor Readings
- Natural Language
- *Almost anything that we can represent as a feature*

$$\mathcal{T} = \{\mathcal{Y}, p(Y|X)\}$$

Task	Label Space	Learned Probability
------	-------------	---------------------

- Object Classification
- Dolphin/Shark Classification
- Sentiment Analysis
- *Any labeled task for which we might be able to build a classifier*



# Transfer Learning

$$X = x_1, x_2, \dots, x_N \in \mathcal{X}$$

$$Y = y_1, y_2, \dots, y_N \in \mathcal{Y}$$

$$\mathcal{D} = \{\mathcal{X}, p(X)\}$$

$$\mathcal{T} = \{\mathcal{Y}, p(Y|X)\}$$

Domain      Feature Space      Probability Observation

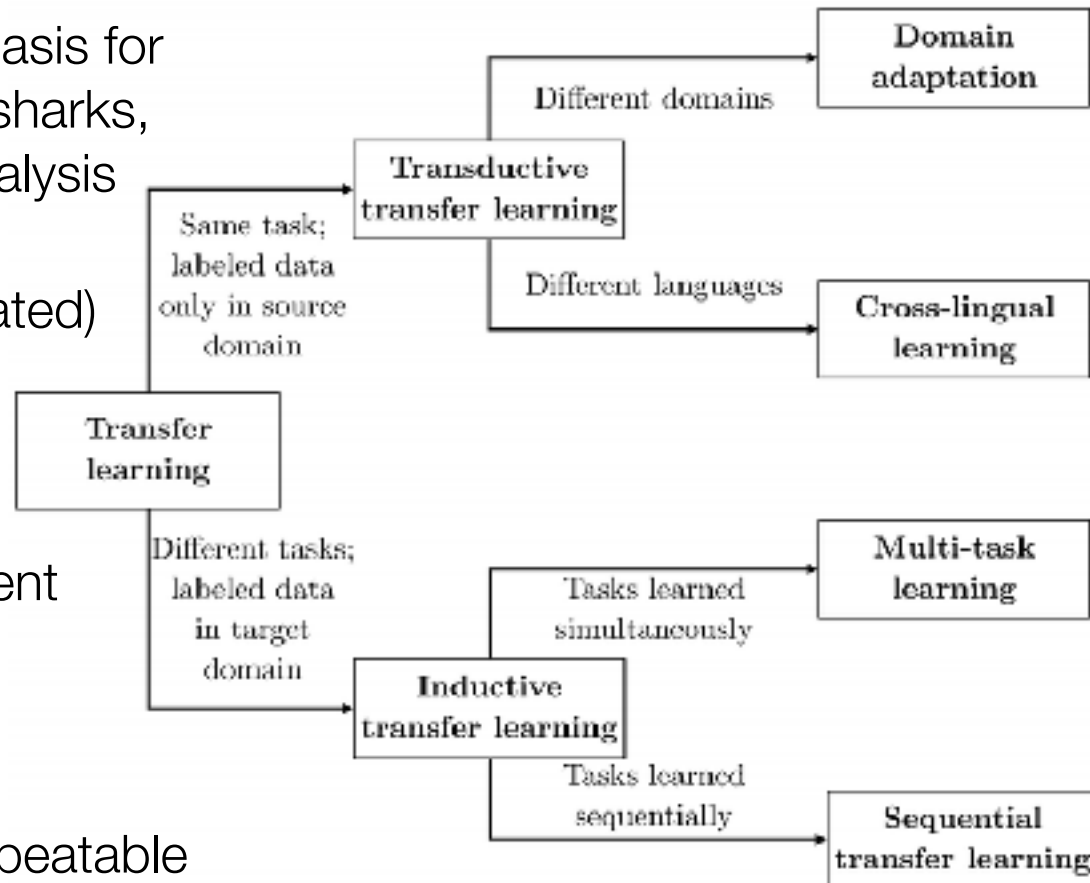
Task      Label Space      Learned Probability

- Need to translate document **Source** to **Target**       $\mathcal{T}_S \rightarrow \mathcal{T}_T$
- Variety of differences might be present. For example, in the context of document classification:
  - **Feature space**: different languages  $\mathcal{X}_S \neq \mathcal{X}_T$
  - **Marginals**: same language, same label space, but differing topics  $p(X_S) \neq p(X_T)$
  - **Conditional**: different label distributions or possibly different labels  $p(Y_S|X_S) \neq p(Y_T|X_T)$



# Categories of Transfer Learning

- **Inductive:** Same Domain, Different Task
  - Using pre-trained VGG as basis for classifying dolphins versus sharks, Style Transfer, sentiment analysis from Glove
- **Transductive:** Different (but related) Domains, Same Task
  - Place identification from RGB Images or LIDAR
- **Unsupervised Transfer:** Different Domains, Different Tasks
  - Learning to paint art and learning to be a surgeon
  - Not yet a field with much repeatable traction



# Other categorizations

	Training	Testing
Transfer Learning	Task 1	Task 2
Multi-task Learning	Task 1 ... Task N	Task 1 ... Task N
Lifelong Learning	Task 1 ... Task N	Task N+1

**Lifelong Learning is a Grand AI Challenge:** Humans can learn to ride a bike and use that to understand better about driving a car. Machine Learning in its current form is far from this capability. How can we move our siloed version of artificial intelligence closer to the process of human based learning? How can we accumulate knowledge from model to model?

Does biology of human learning hold any clues to success? How does a human learn to crawl? To talk? To ride a bike? What is a human's motivation to learn?

