Encoding human priors: Data augmentation and prompt engineering

Sharon Zhou

January 26, 2023

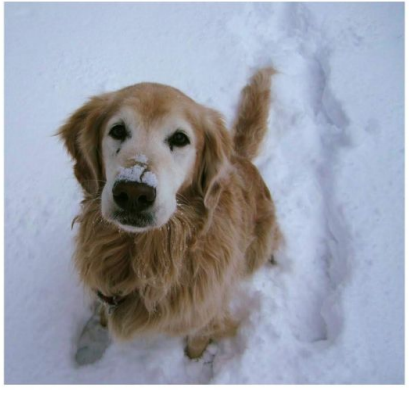
Contents

- ML models fail in simple ways
  - Encoding human priors into data
  - Overfitting/underfitting
- Human priors to augment training data
  - Data augmentation
  - Types of augmentation
- Human priors at test-time (LLMs)
  - Prompt engineering
  - Context matters
  - Correcting input data

ML Models Fail in Simple Ways

#fail

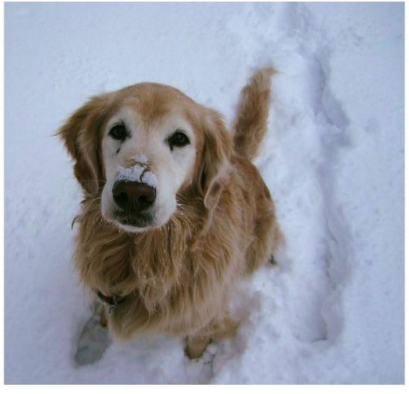
Model trained on dogs like this...



thinks this is a cat.

This looks ridiculous

Model trained on dogs like this...



thinks this is a cat.

But that’s because of our priors

Human priors

Knowledge about the data and task

(That we often take for granted)

e.g. a rotated picture of a dog still has a dog

Priors over invariances are useful

Invariances

Changes to the input that don’t change its output

e.g. a rotated picture of a dog still has a dog

Encoding human priors into data

Encode = find a function to represent

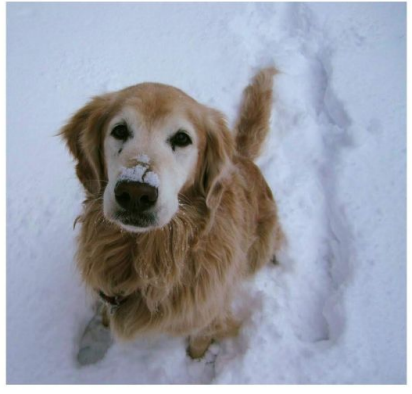
human priors = our knowledge about the task

into data = that operates on the data

Result: transforms on the input data

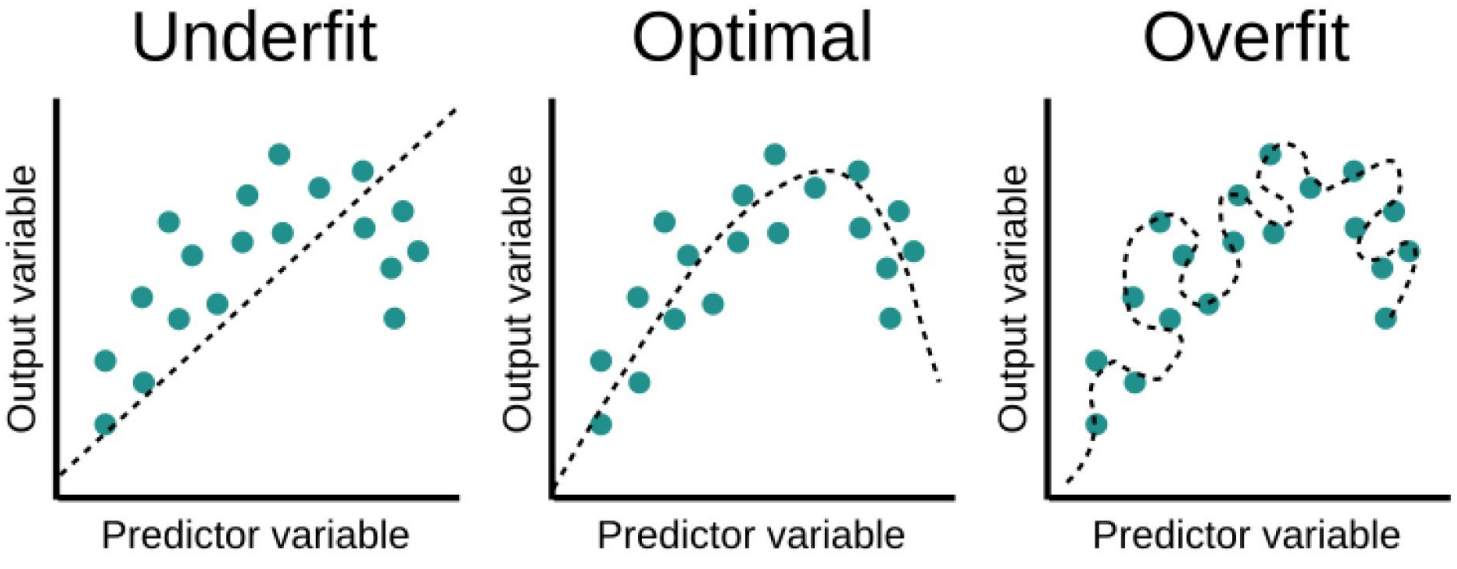
#win

Model trained on dogs like this...



thinks this is a dog too.

In machine **learning** terms



Human Priors to Augment Training Data

Collecting enough data is hard

Data augmentation

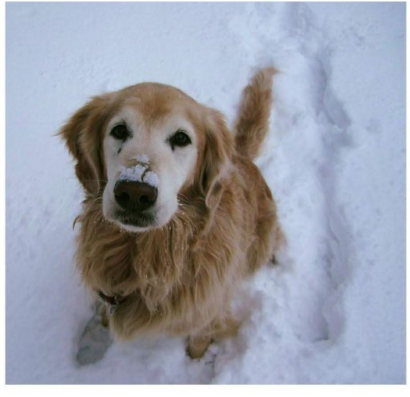
Add more data with the data you already have.

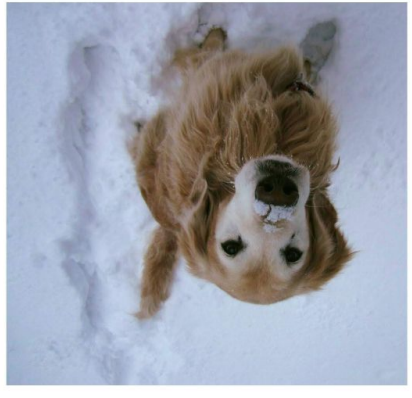
Easy win to improve your model without more data.

Encode human priors over invariances in your data.

Data augmentation for images

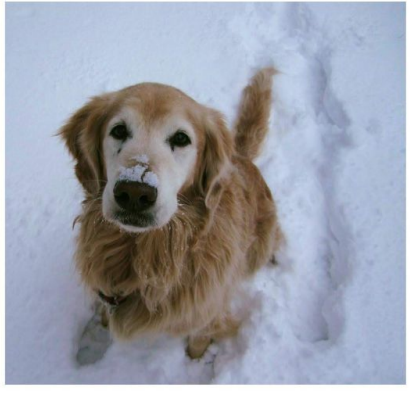
Flip and rotation

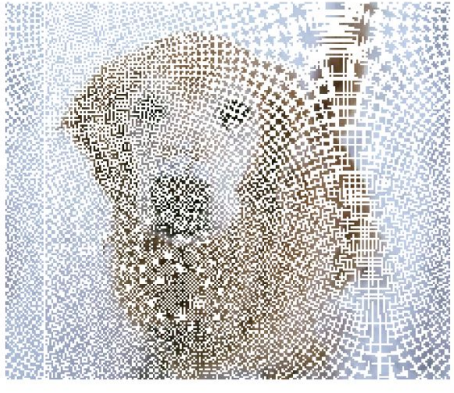




Data augmentation for images

Mobius transformations





Data augmentation with Mobius transformations, https://arxiv.org/abs/2002.02917

Data augmentation for images

MixUp

60% cat

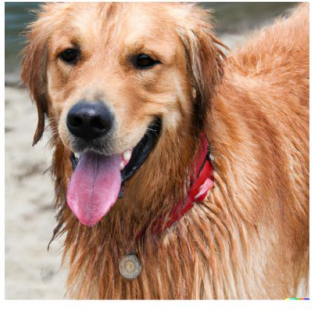
40% dog

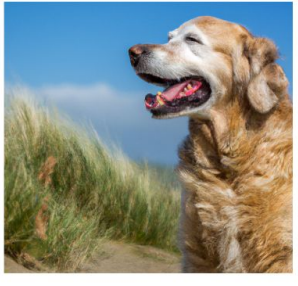
mixup: Beyond Empirical Risk Minimization, https://doi.org/10.48550/arxiv.1710.09412

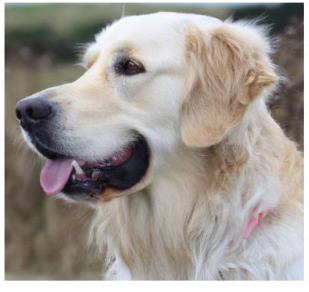
Data augmentation for images

Synthetic data generation

Real







Generated
by DALL-E 2

Data augmentation for images

Sim-to-real transfer for robotics

RetinaGAN: An Object-aware Approach to Sim-to-Real Transfer

Data augmentation for text

Back-translation

Human Priors at Test-Time (LLMs)

Prompt engineering

Change the input at test-time to elicit desired results

Write a letter of recommendation for a student.

Write a letter of recommendation for a student who gets into MIT.

LLMs have an easy language interface

Language is easy for people to engineer prompts

LLMs have an easy language interface

Language is easy for people to engineer prompts

Other modalities have knobs too, but language provides an easier interface

(e.g. image disentanglement)

unicorn llama by Stable Diffusion

Prompt engineering depends on the model

ChatGPT is finetuned to accept commands better

GPT-3

What is your first name?

What is your last name?

GPT-3.5 (ChatGPT)

What is your first name?

My first name is Sarah.

Giving examples in the prompt

An example can help nudge what you want

GPT-3

What is your last name?
Kane.

What is your first name?
Tom.

Lab: Prompt Engineering

Explore changes in the context (aka. prompt) (aka. data).

Build a context template for scalability and reusability.

Add examples to boost performance.

Conclusion? *Change the data. Change everything.*

Encoding human priors: Data augmentation and prompt engineering

Sharon Zhou

January 26, 2023