Introduction to Statistical learning

Omid Safarzadeh

January 18, 2022

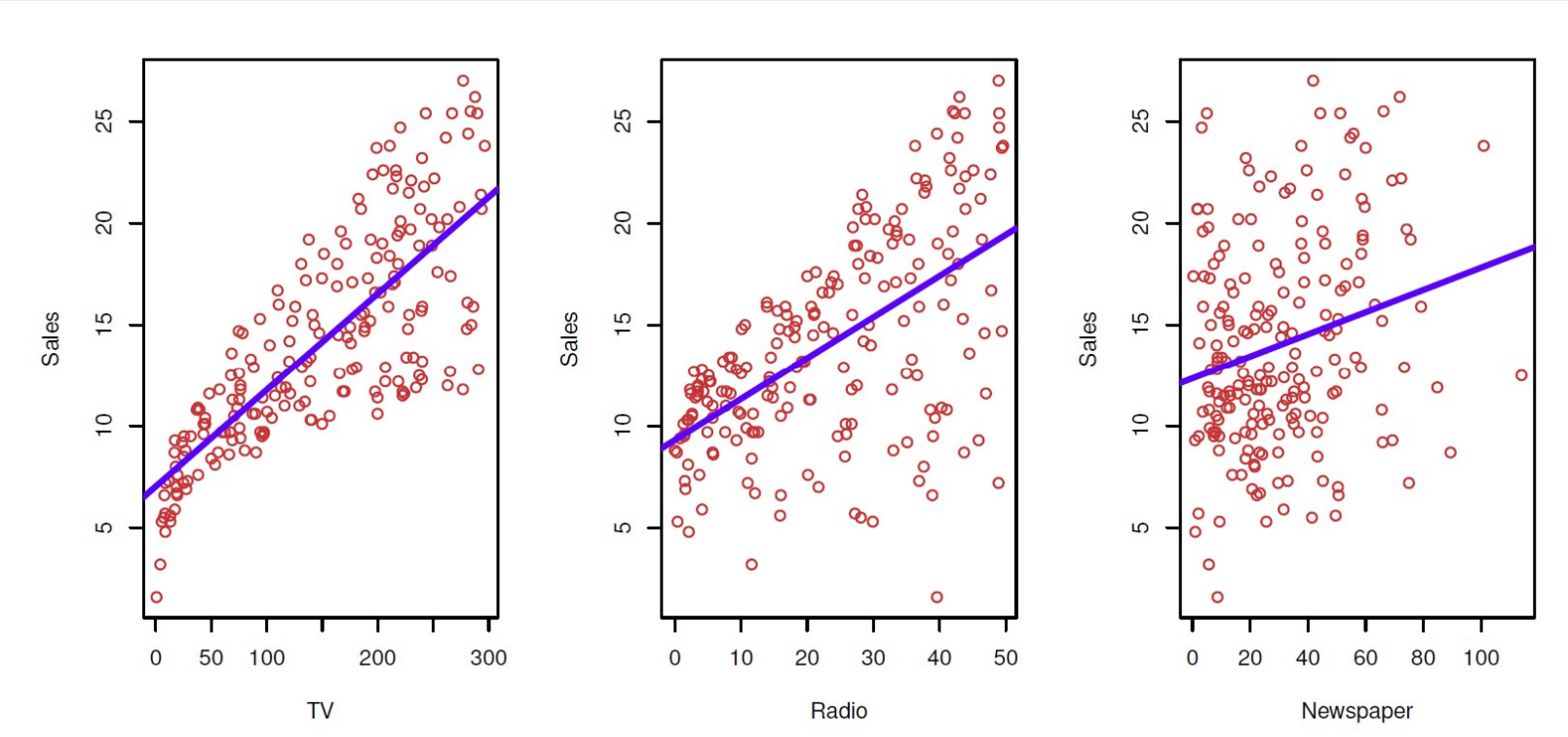
Table of contents

Supervised learning

Model: $Y_{output} = f(X_{input}) + \epsilon_{noise}$ **Training data:** $(x_{i\,instance}, y_{i\,label}), i = 1, ..., n.$ **Learning process:** $\{(x_i, y_i)\}_{i=1}^{n} \rightarrow$Learning algorithm $\rightarrow \hat{f}(.)$ **Goal:** Ensure that $\hat{f}(X)$ is close to Y for all possible X and Y pairs.

Example 1.1

Advertising: X: Advertising budget (thousands of dollars). Y: Sales (thousands of units). Each dot corresponds to a previous advertising campaign.



Blue line: least squares estimate of sales given data.

Example 1.2

CIFAR 10:

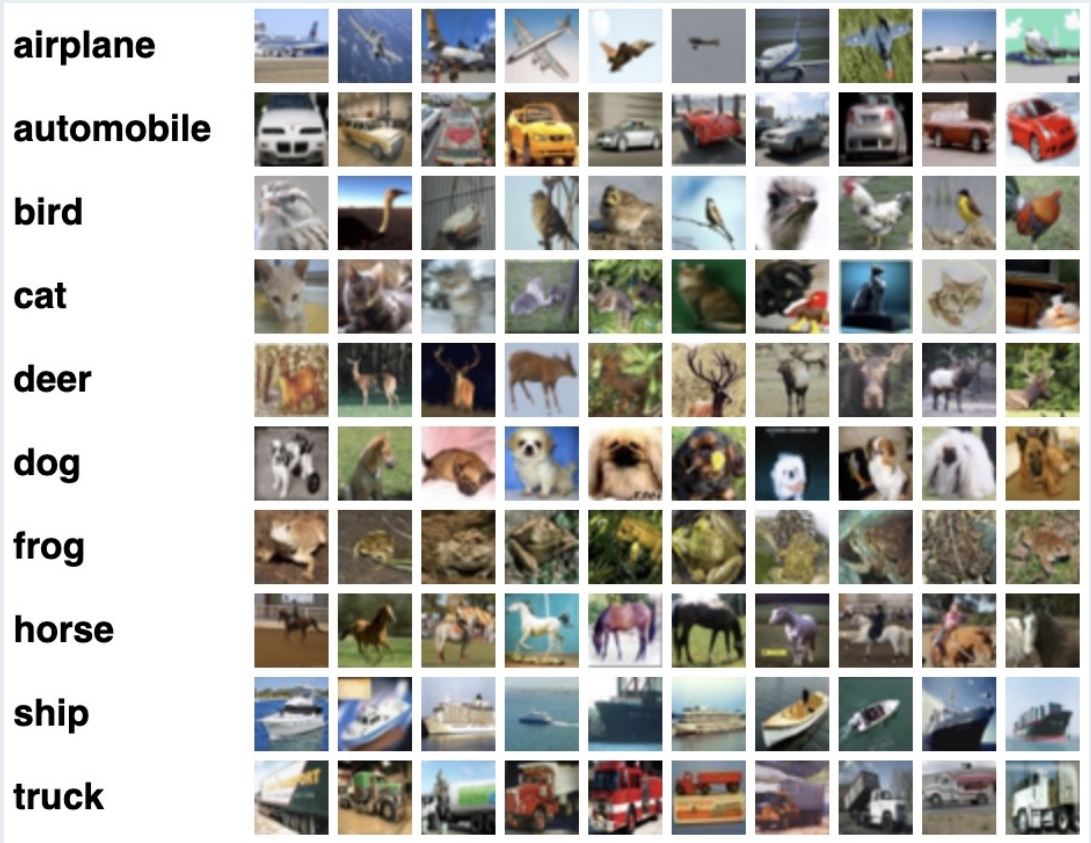
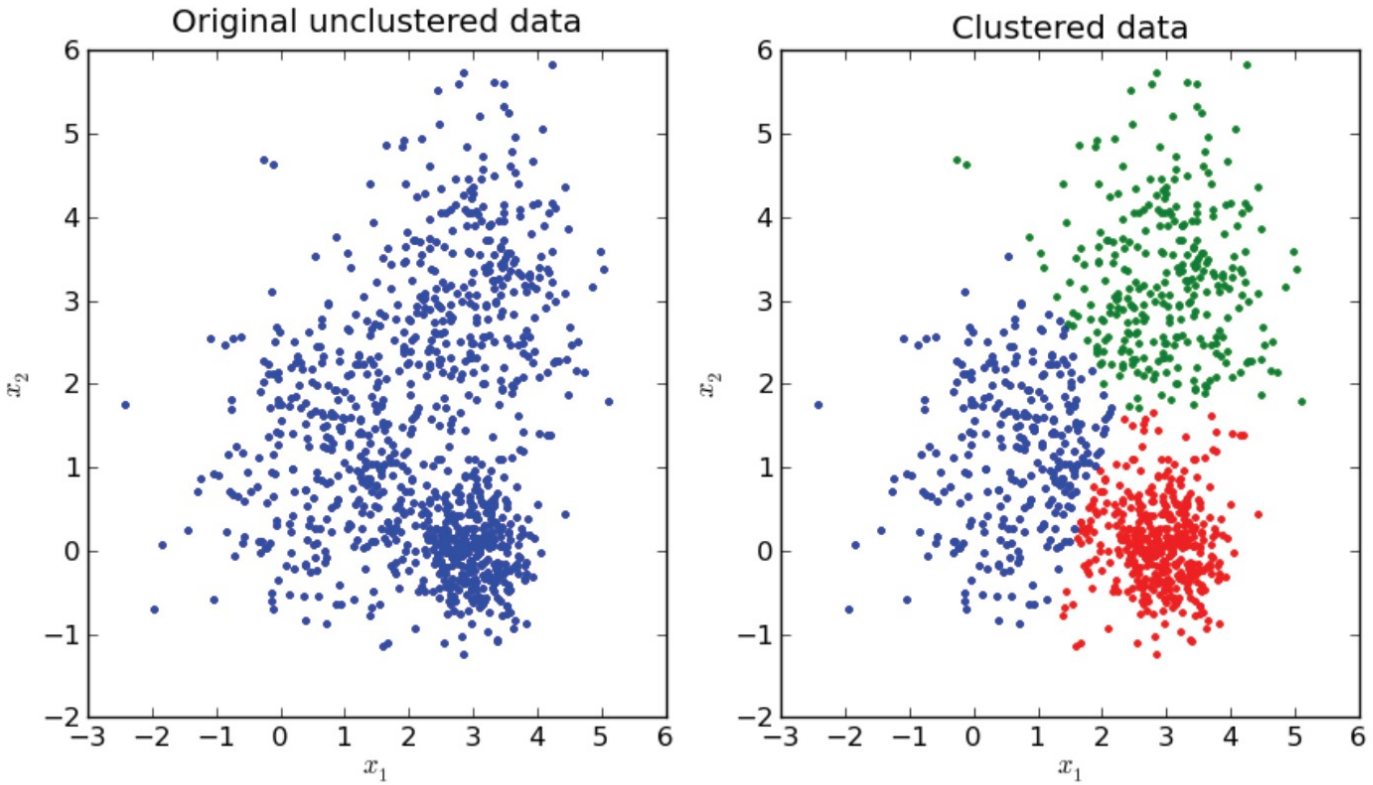


Figure: Krizhevsky, Hinton, et al., 2009

Unsupervised learning

Training data: $x_i, i = 1, ..., n$. (no labels) Can we recognise data into different groups?



K-means with $K = 3$ clusters.

Example 2.1

Image Compression Image compression refers to the task of representing images using as little storage (i.e., bits) as possible.

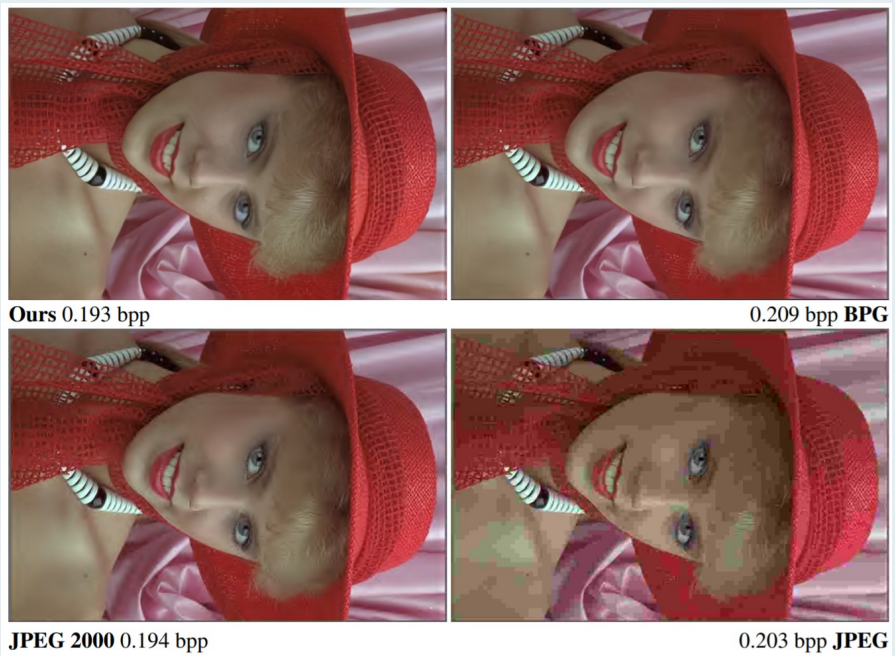


Figure: Conditional Probability approach vs. BPG, JPEG and JPEG 2000 on the third and fourth image of the Kodak data set. Mentzer et al., 2018

Example 2.2

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

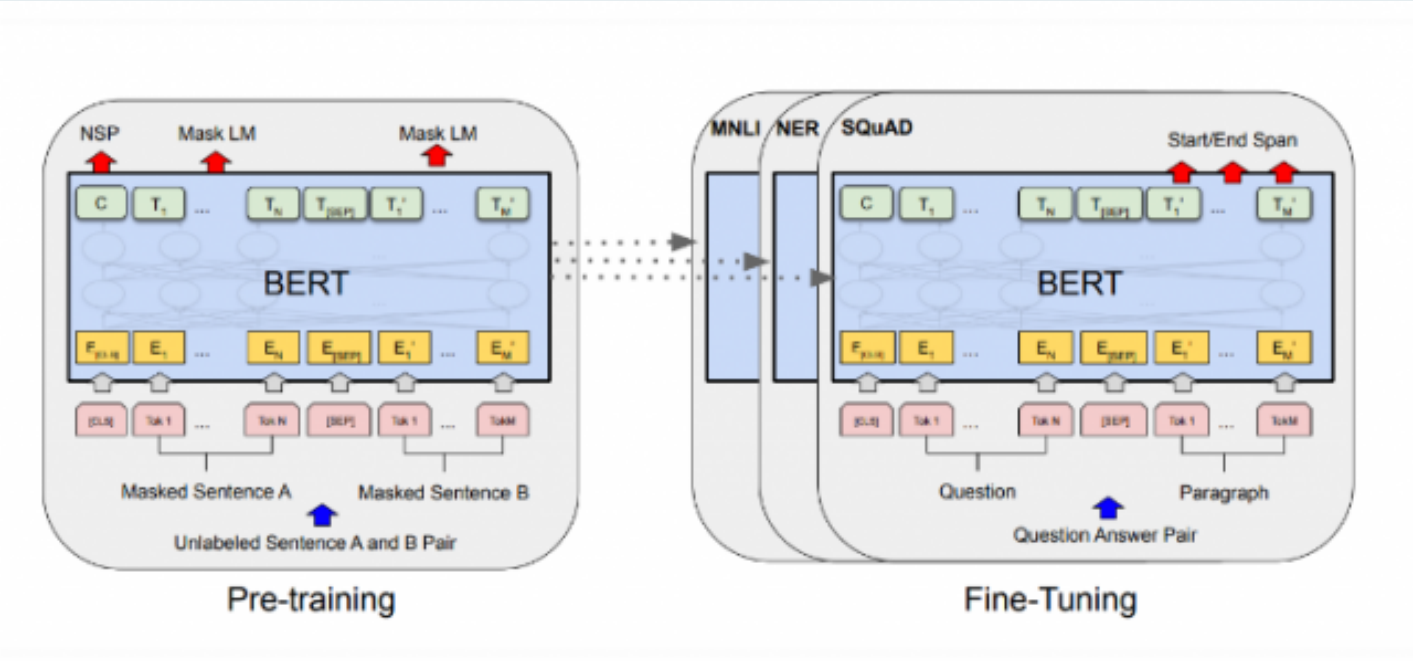


Figure: Devlin et al., 2018

Semi-Supervised learning

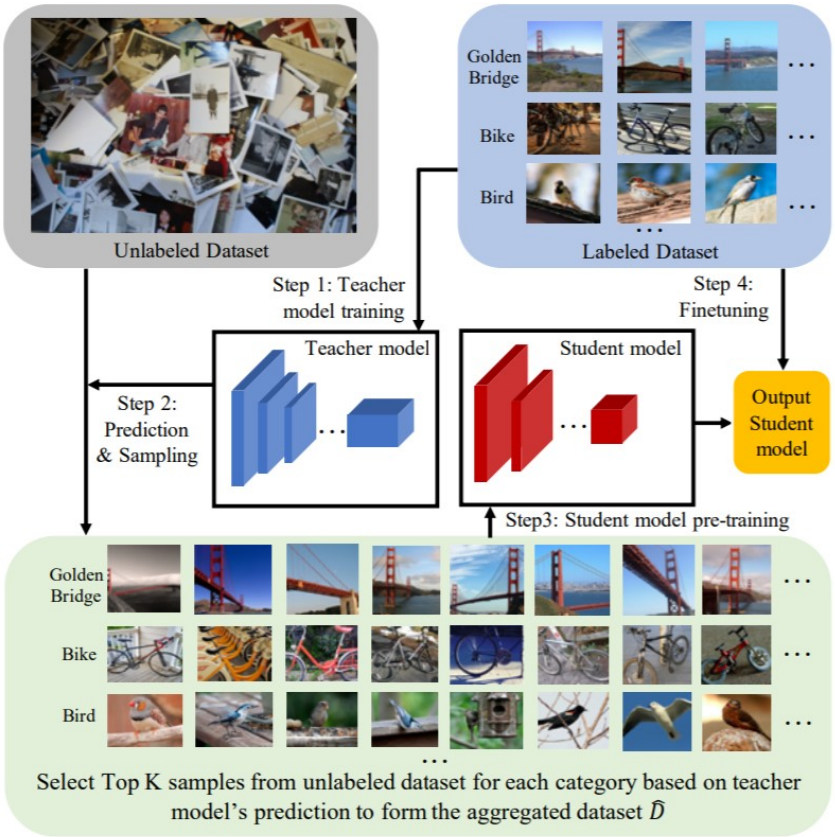


Figure 1: Illustration of our approach: with a strong teacher model, we extract from a very large unlabelled image collection (100M–1 billion images) a new (large) training set. The student model is first trained with this noisy supervision, and fine-tuned with the original dataset.

Figure: Yalniz et al., 2019

Some examples

- **Deep learning - Deep belief network**:
  Handwritten digit classification and generation
  Example: Link
- **Reinforcement learning**
  1. **Hide and seek**
     Example: Link
  2. **Alphago and Alphazero**
     Example: Alphago Alphazero

Online Learning

supervised/unsupervised learning:

- feeding data in batch to model
- data set is static
- not useful for streaming data
- the models become outdated after a while
- Concept drift?

Solution: incremental or online learning algorithms.
REF:Putatunda, 2021

Reinforcement Learning

Graph Representation Learning

References

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Krizhevsky, A., Hinton, G. et al. (2009). Learning multiple layers of features from tiny images.
- Mentzer, F., Agustsson, E., Tschannen, M., Timofte, R., & Van Gool, L. (2018). Conditional probability models for deep image compression. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4394–4402.
- Putatunda, S. (2021). Practical machine learning for streaming data with python.
- Yalniz, I. Z., Jégou, H., Chen, K., Paluri, M., & Mahajan, D. (2019). Billion-scale semi-supervised learning for image classification. *arXiv preprint arXiv:1905.00546*.