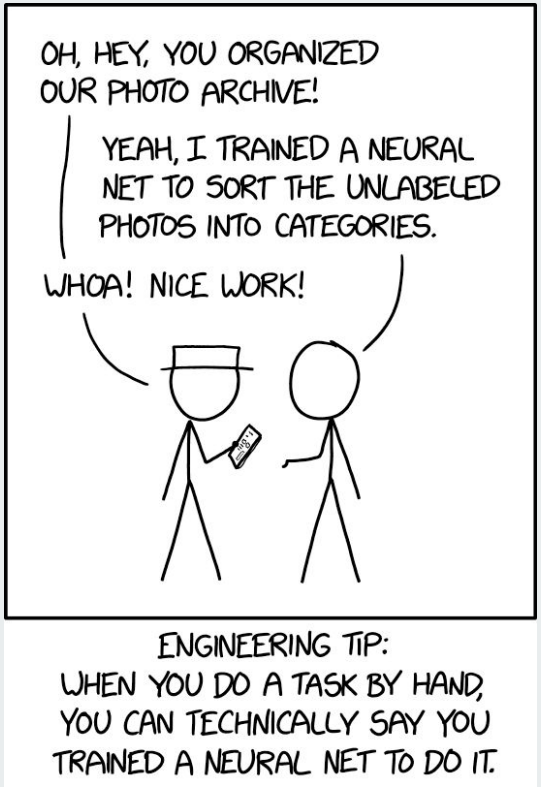
https://www.explainxkcd.com/wiki/index.php/2173:_Trained_a_Neural_Net

Deep Learning Basics & Recap

Cassie Guo



http://www.deeplearningbook.org/contents/intro.html
https://vwo.com/ab-testing-openai-gpt-3/results/?utm_medium=social&utm_source=twitter&utm_campaign=tof_eg_humansvsai

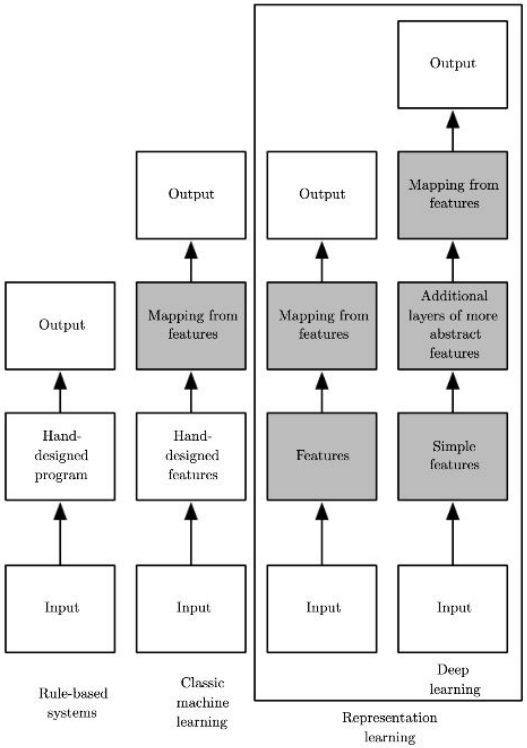
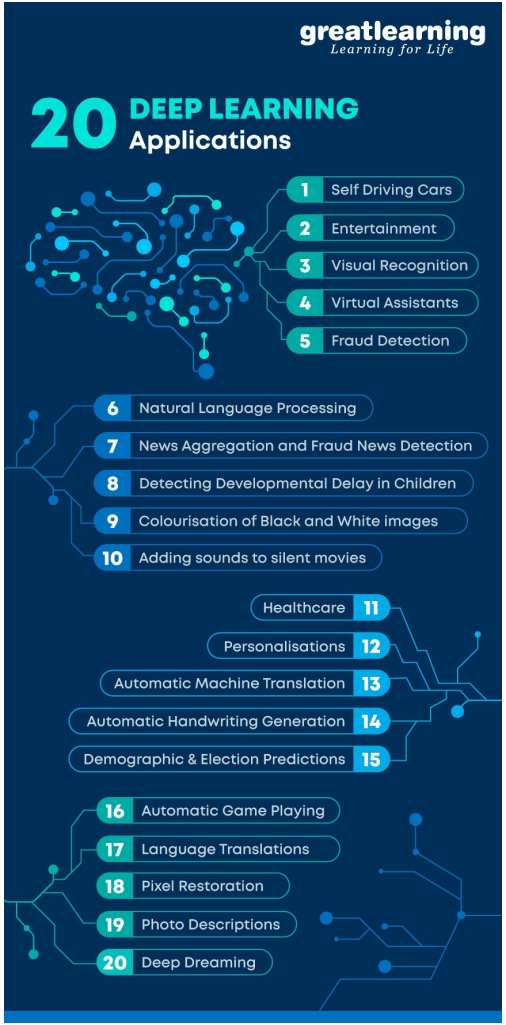
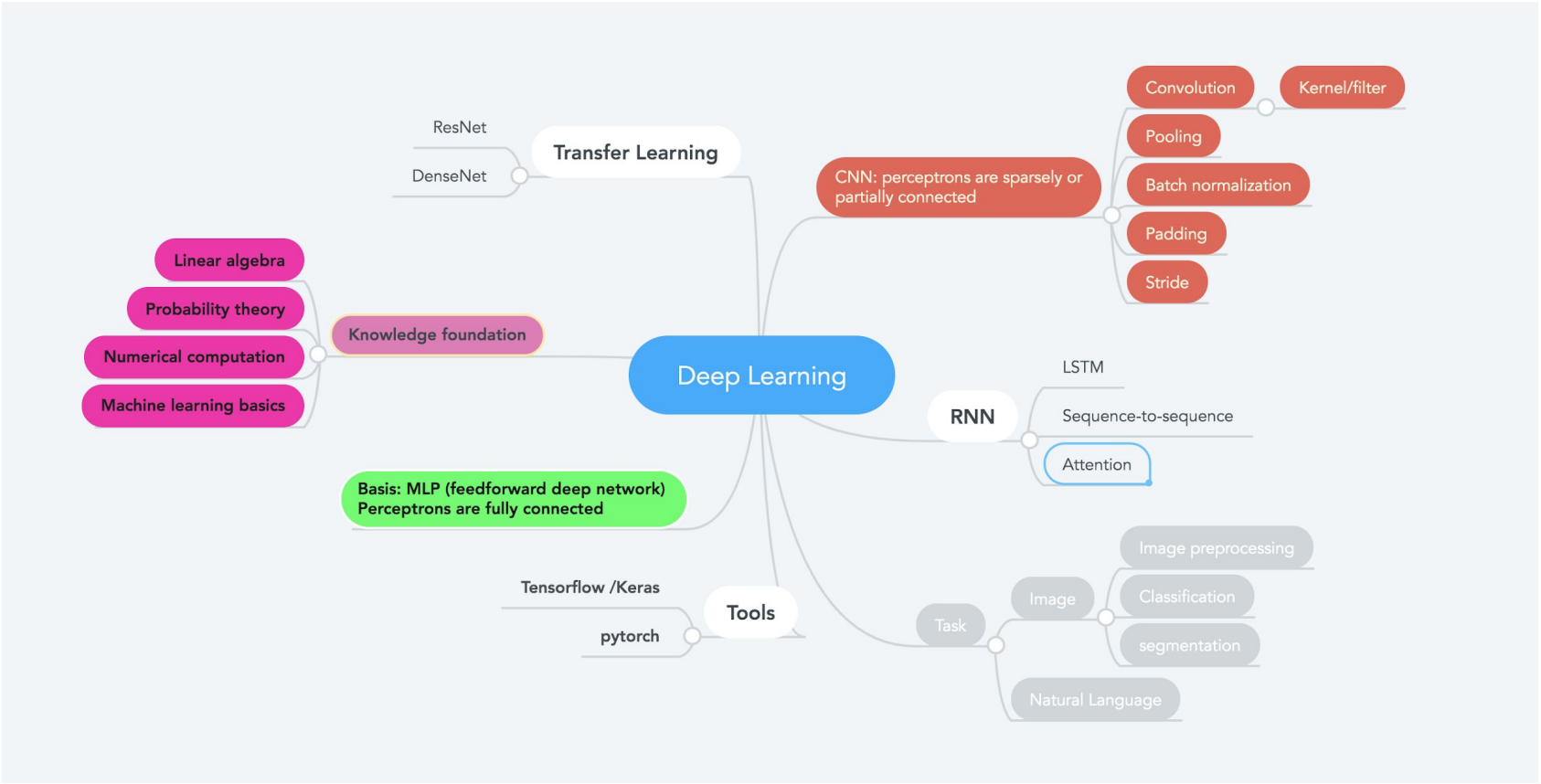


Figure 1.5: Flowcharts showing how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data.





Knowledge Map

Resources used

- MIT Deep Learning
- Fast.ai
- Kaggle minicourse
- Deeplearning.ai
- Others

Foundations

- Part I: Applied Math and Machine Learning Basics
  - **2 Linear Algebra**
    - Tensors, matrix multiplication, dot product, SVD, norm
  - **3 Probability and Information Theory**
    - Conditional probability chain rule, bayes rule, gaussian, binomial
  - **4 Numerical Computation**
    - Underflow and overflow, Jacobian & Hessian, gradient-based optimization
  - **5 Machine Learning Basics**
    - Cross validation, bias & variance

Deep Feedforward Networks

- 6.1 Example: learning XOR
- 6.2 Gradient-based learning
  - Cost function (negative log likelihood, maximum likelihood, cross-entropy)
  - Output units (linear, sigmoid, softmax)
- 6.3 Hidden units
  - Relu
  - Logistic sigmoid and Tanh
  - Others (RBS, softplus, hard Tanh)
- 6.4 Architecture design
- 6.5 Backpropagation and other differentiation algorithms
  - Intuition
  - Notes

Regularization in Deep Learning

- 7.1 Parameter Norm Penalties (L1, L2) (weight decay)
- 7.2 Norm Penalties as Constrained Optimization
- 7.3 Regularization and Under-Constrained Problems
- 7.4 Dataset Augmentation (Pytorch example)
- 7.5 Noise Robustness (Label smoothing)
- 7.6 Semi-supervised Learning
- 7.7 Multitask Learning

Regularization in Deep Learning

- 7.8 Early Stopping
- 7.9 Parameter Tuning and Parameter Sharing (CNN)
- 7.10 Sparse Representations
- 7.11 Bagging and Other Ensemble Methods
- 7.12 Dropout
- 7.13 Adversarial Training
- 7.14 Tangent Distance, Tangent Prop and Manifold Tangent Classifier

Optimization for Training Deep Models

- 8.1 How Learnings Differs from Pure Optimization
- 8.2 Challenges in Neural Network Optimization (ill-conditioning, local minima, saddle point, etc)
- 8.3 Basic Algorithms (SGD, Momentum, Nesterov Momentum)
- 8.4 Parameter Initialization Strategies
- 8.5 Algorithms with Adaptive Learning Rates (AdaGrad, RMSProp, Adam)
- 8.6 Approximate Second-Order Methods (Newton's Method, BFGS, etc)
- 8.7 Optimize Strategies and Meta-algorithm (Batch Normalization, Coordinate Descent, Polyak Averaging, Supervised Pretraining, etc)

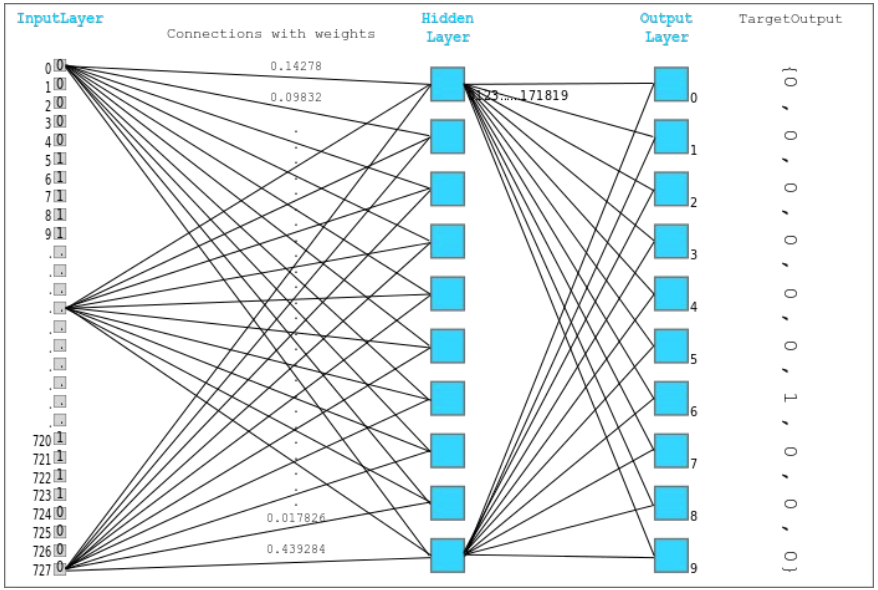
Experiential Learning: how to learn deep learning

https://en.wikipedia.org/wiki/David_A._Kolb

Example of Related Interview Questions

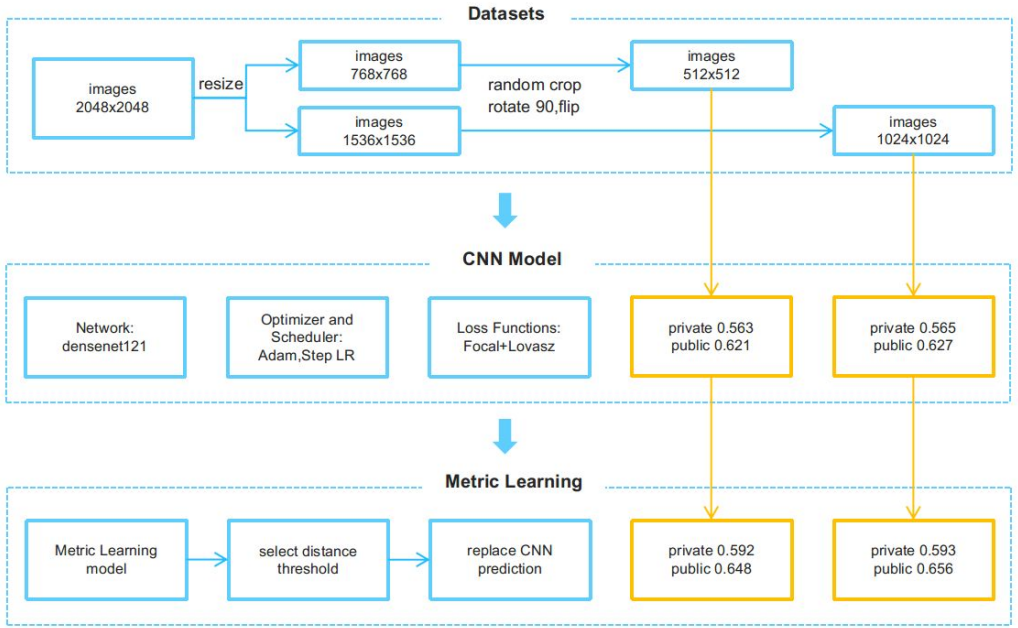
- Explain gradient vanishing/explosion
- What is back propagation
- What is attention layer in RNN
- Why convolution works
- Dimension calculation: given input size, padding size, kernel size, calculating the output size
- What kind of data augmentation
- Why do we need batch normalization

Build it from scratch



- What methods do I need for class MnistModel()
- Constructor?
- What are the goals of each function?
- What are the inputs, outputs?

Kaggle Example



- Image preprocessing
- DenseNet architecture
- Learning rate & scheduler?
- Loss function?
- Adaptive Max Pooling?
- Relu?
- Why and how
- Other suboptimal sulutions

Useful resources

- 群内之前活动slides by Xianjun
- Book
  - MIT Deep Learning
  - Approaching (Almost) Any Machine Learning Problem
  - Elements of Statistical Learning
  - ?
- Course
  - Fast.ai
  - Kaggle minicourse
  - Deeplearning.ai
  - ZeroToGans
  -
  - ?

Useful resources

- Videos
  - Statsquest
  - Hong-yi Lee
  - AlphaOpt
  - KhanAcademy
- Papers
  - Paper with Code
  - Distill
- Others
  - Kaggle Winning Solutions