

## 第75次活动: 09/18/2020

Saturday, April 18, 2020 7:03 PM

现场人员出席状况	<a href="http://bit.ly/ml_names">http://bit.ly/ml_names</a>
时间	09/18/2020, 周五 8PM - 10PM
地点	Online
Remote dial-in	<a href="https://zoom.us/j/115286882">https://zoom.us/j/115286882</a>
活动主题	Deep Learning 基础知识大纲
Lead discussion	Cassie Guo

录像: [youtube link](#)  
[Slides](#)



Deep  
Learning

原始视频, 请勿外传  
<https://1drv.ms/v/s!Avt033d3-E4ahrX0WSI3H9EyPoL6sg?e=2i656y>

From Eric-mlp to Everyone: 08:42 PM

这个dimension比数据多  
 然后matrix不能invert  
 加入regularization问题就可解了

From Ray to Everyone: 08:48 PM

A Visual Survey of Data Augmentation in NLP: <https://amitnss.com/2020/05/data-augmentation-for-nlp/>

From Jiaying Shi to Everyone: 09:11 PM

[http://videolectures.net/deeplearning2015\\_montreal/](http://videolectures.net/deeplearning2015_montreal/) 这个会议比较早, 里面有一些关于理论的讨论, ian goodfellow 有两个talk 内容都很有趣

From Xinyi to Everyone: 09:15 PM

Pinterest的那个multi task embedding paper能分享一下吗?

From Guocui Mi to Everyone: 09:16 PM

[https://labs.pinterest.com/user/themes/pin\\_labs/assets/paper/learningunifiedembedding-kdd2019.pdf](https://labs.pinterest.com/user/themes/pin_labs/assets/paper/learningunifiedembedding-kdd2019.pdf)

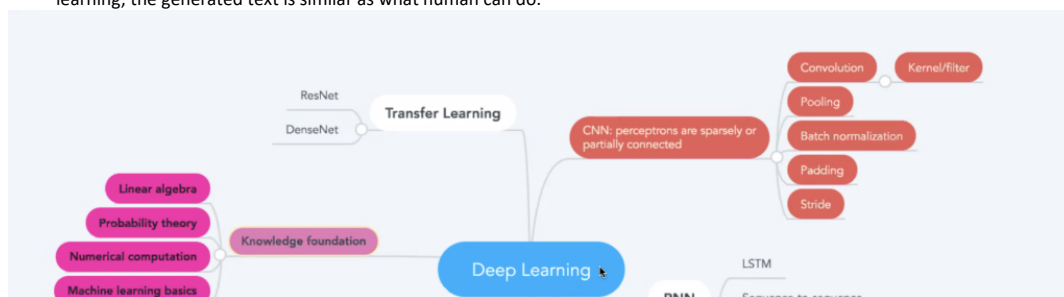
Logistics from Xianjun: 目前focus在DP上

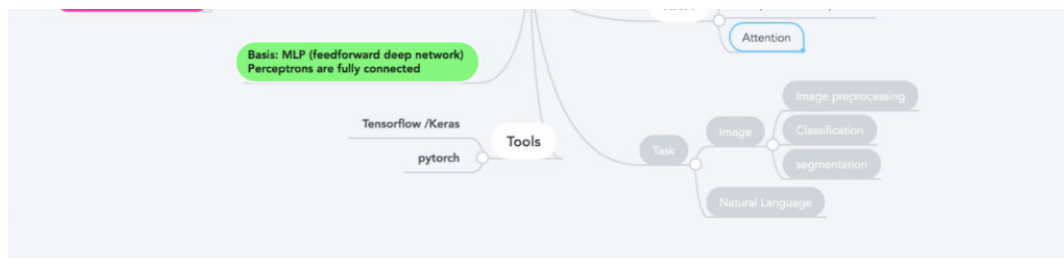
Slide link:

[https://docs.google.com/presentation/d/1wufYSW22SPvtBdUkY8dRqIPigqMP1SobjO9bZahiJ5k/edit#slide=id.g968825bfcd\\_0\\_47](https://docs.google.com/presentation/d/1wufYSW22SPvtBdUkY8dRqIPigqMP1SobjO9bZahiJ5k/edit#slide=id.g968825bfcd_0_47)

Cassie:

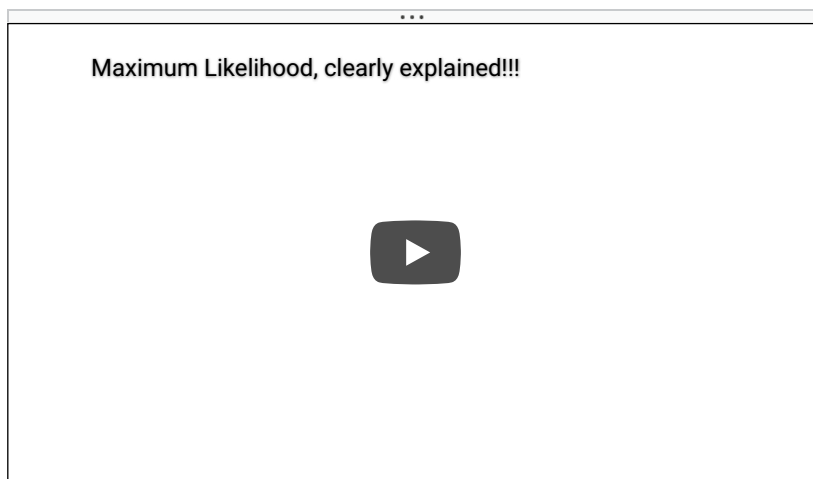
- Not much involved in work using DP; use Kaggle a lot
- The content in the slide is mainly based on the flower book (DP)
- Lots of application of DP: it has many layers and can extract abstract features, while comparing with traditional ML, traditional ML has lots of feature engineering need to be done.
- GPT-3 VS human
  - Demoed a A/B test use case using GPT-3
  - What is GPT-3? it learns from lots of data to do text generation, using unsupervised learning; the generated text is similar as what human can do.





## Knowledge Map

- Not very familiar yet with the actual optimization algorithms, interested in learning more about it.
- To learn DP, understanding how the optimization is done, and also how tensors work can be very helpful
  - Tensor is basic in deep learning, it can be number, vector, matrix, or n-dimension
- Possible interview content: SVD – PCA steps, bayes rules
- **All the links embedded in the links on the slides are highly recommended resources for learning**
- XOR, how to develop a Deep forward network, multi-layer, basic neural network. Different than CNN (partially connected), XOR is totally all connected.
  - Highly recommend this video to understand maximum likelihood: [StatQuest: Maximum Likelihood, clearly explained!!!](#)



- Backpropagation -- to calculate the gradient coefficients.
- Different regularization methods, L1 (lasso, Laplace prior) and L2 (ridge, gaussian prior, weight decay) are the basics, how to add the penalty to the loss function
- <<elements of statistical learning>> book recommend, we'll explain about L1 and L2
- Norm penalties as constrained optimization 相当于 L1 and L2
- Dataset augmentation in image 很常用, 数量有限, 变换照片, 平移, 剪切
- Wenxiang and others:
  - 数量很大的时候, 对比 contrast learning, transfer learning
  - text 方面, 没有很好的 data augmentation, BERT, GPT3 都是 unsupervised learning
  - Text augmentation, back translation 是一种 technique example
- label smoothing
  - Yuhang, Tienan have experience
  - What is label smoothing? 假如 10 个 class, 每次预测一个 class, 传统的用 1 和 0 来 label, label smoothing 用 0~1 间, 比如 0.8 来 label, 让学的过程 更加 robust, regularized; Penalty 变弱, 切入点在 label 上。
  - 应用例子, sentiment analysis, 图像识别等 和 抽象概念有关的, 有认知不统一或者偏差的情况下
  - Jason Xu: CLR, click through 的 quality 不一样, 10 分钟才 click 和 1 分钟 click, 1 分钟的可能是一个 false click
  - 当没有 confident 的时候, 来去 compromise, 对 label 进行处理, 利用其它 surrogate 的信息来衡量你的 target 的准确度
- Semi supervised learning
  - Wenxiang: 只用 label 一部分, 来学习预测其它未 label 的数据; 两部分的 Loss 不一样; unsupervised and supervised 同时
  - Yuhang: fraud detection example, 2% has clean label, 98% has dirty label
  - SSL, Google 的 UDA 类似方法。Snokel --> 效果有限
  - 现在流行的方法有:
    - Self-learning
    - Contrasting learning
      - Image --> augmentation, 预训练, 和 BERT 类似
    - Active learning: self-driving; 发现有价值的东西, 一直把东西, 纠正 false 的信息
  - Semi learning:
    - TEXT data --> BERT
    - Image -->

- Structure data 不太确定如何处理。
- Comments: Deep learning 像炼丹, data validation 后, 解决某一个具体的问题
- Multitask learning
  - tesla, CNN, backbone, image 提取feature, 识别红绿灯, 同时的训练, tweak同时在训练backbone,
  - Eileen 例子: search 工作, 找相关性, perfect match, 还是bad。假如 5 个classes, click 还是没有, 或者 top 10? 先学randomness, predict, loss function 包含多个部分, 小数据不work, 要大数据, 长时间训练。有时间, 限制条件, train的时间要足够长,
  - Multi-task, 要balance 几个不同的tasks, loss 的weight 来调整, 如何control loss, 比如通过sample size, Loss function 可以不一样,
  - Fine tune -- 》1 层的layer 就好了, 所有的任务进来 用同一个embedding; 所有Tasks 同时先fine tune 一个backbone, 再加最后一层根据不同的goal。
- Dropout 以前常用, 但现在不常用, batch normalization (look into distribution, while dropout breaks the distribution)冲突, 所以后来不常用
- Adaptive grad learning: How to schedule learning rate, so to improve the performance, one cycle learning?
- **Experimental learning framework:**
  - Jovian.ml, 4 chapters, pytorch, 3 hours, cover
  - FAST.AI, solve problem through code
  - 看别人的kaggle kernel来学习
- Cassie背景: travel industry before, now work in start up; did fraud, recommendation, NLP 等;