

李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

中英双语字幕



课件

一键打包下载



笔记

官方笔记翻译



代码

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1fM4y137M4>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/ntu-hylee-ml>

机器学习
深度学习

批次标准化

Auto-encoder

卷积神经网络

生成式对抗网络

GAN

自监督

强化学习

元学习

学习率

自注意力机制

Transformer

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



微信公众号

资料下载方式 2: 扫码点击**底部菜单栏**
称为 **AI 内容创作者**? 回复 [添砖加瓦]

Machine Learning HW9

Explainable AI

ML TAs

ntu-ml-2021spring-ta@googlegroups.com

Outline

- Topic I: CNN
 - Model & dataset
 - Task
 - Lime
 - Saliency Map
 - Smooth Grad
 - Filter Visualization
 - Integrated Gradient
- Topic II: BERT
 - Attention Visualization
 - Embedding Visualization
 - Embedding Analysis

Topic I: CNN explanation

Model: food classification

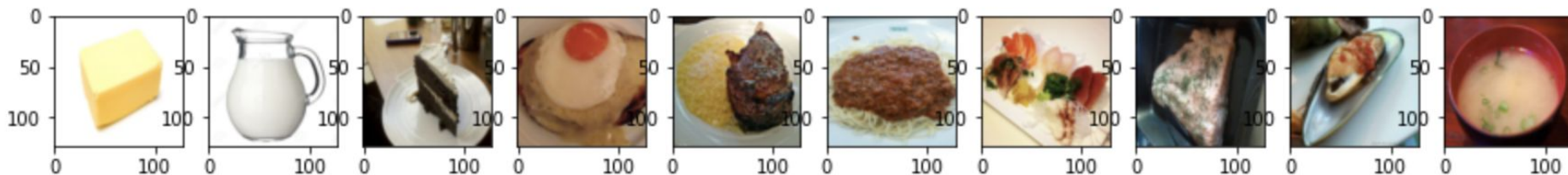
- We use a trained classifier model to do some explanations
- The classifier model is a CNN model, aim to classify different kinds of food
- Dataset: 11 categories of food (same dataset in HW3)
- Bread, Dairy product, Dessert, Egg, Fried food, Meat, Noodles/Pasta, Rice, Seafood, Soup, and Vegetable/Fruit
- We only pick up 10 images in trainset for observation

Task

- Run the sample code and finish 20 questions (all multiple choice form)
- We'll cover 5 explanation approaches
 - Lime package
 - Saliency map
 - Smooth Grad
 - Filter Visualization
 - Integrated Gradients
- You need to:
 - Know the basic idea of each method
 - Run the code and observe the results
 - For some case you may need to modify a little part of the code

Task: observation

- To finish this homework, you only need to observe these ten images.
- Please make sure you got these 10 images in your code.
- We encourage you to observe other images!



Lime

Question 1 to 4

- Install the Lime package > `pip install lime==0.1.1.37`

GitHub Repo: <https://github.com/marcotcr/lime>

Ref: <https://goo.gl/anaxvD>

Saliency Map

Question 5 to 9

- Compute the gradient of output category with respect to input image.

Ref:

<https://medium.com/datadriveninvestor/visualizing-neural-networks-using-saliency-maps-in-pytorch-289d8e244ab4>

Smooth Grad

Question 10 to 13

- Randomly add noise to the input image, and get the heatmap. Just like what we did in the saliency method.

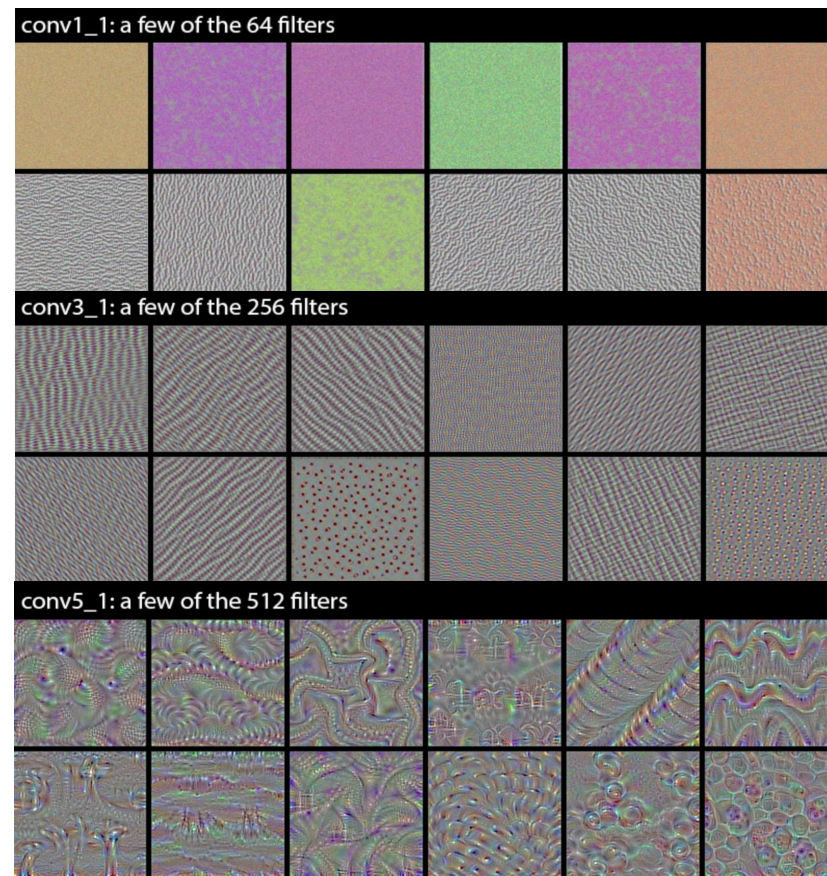
Ref:

<https://arxiv.org/pdf/1706.03825.pdf>

Filter Visualization

Question 14 to 17

- Use **Gradient Ascent** method to find the image that activates the selected filter the most and plot them (start from white noise).



Integrated Gradients

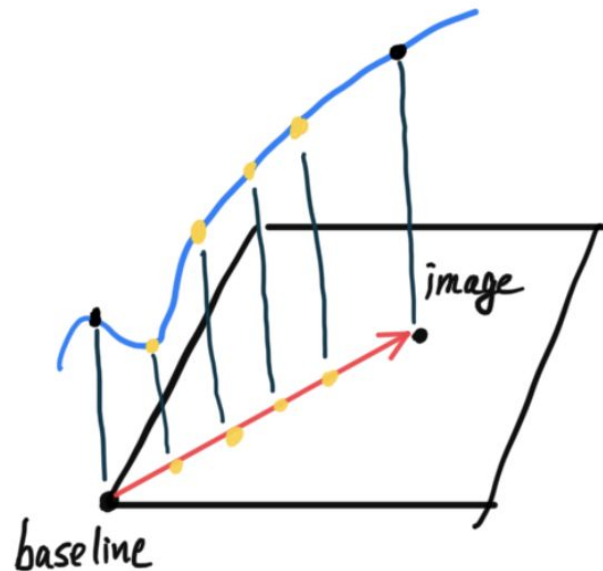
Question 18 to 20

- Flexible baseline

$$(x_i - \bar{x}_i) \cdot \int_{\alpha=0}^1 \frac{\partial S_c(\tilde{x})}{\partial(\tilde{x}_i)} \bigg|_{\tilde{x}=\bar{x}+\alpha(x-\bar{x})} d\alpha$$

Ref:

<https://arxiv.org/pdf/1703.01365.pdf>



Topic II: BERT explanation

Attention Visualization

Question 21 to 24

Visualize attention mechanism of bert using

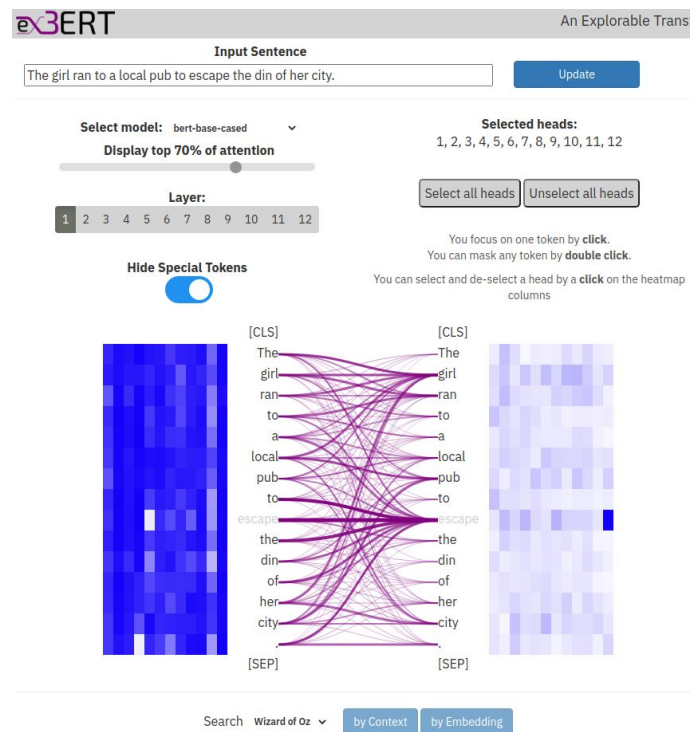
<https://exbert.net/exBERT.html>

Objective:

- (1) What are the functions of different attention heads?
- (2) How does the model predict masked words?

Alternative Link

<https://huggingface.co/exbert>



Paper: <https://arxiv.org/abs/1910.05276>

Tutorial: https://youtu.be/e31oyfo_thY

Embedding Visualization

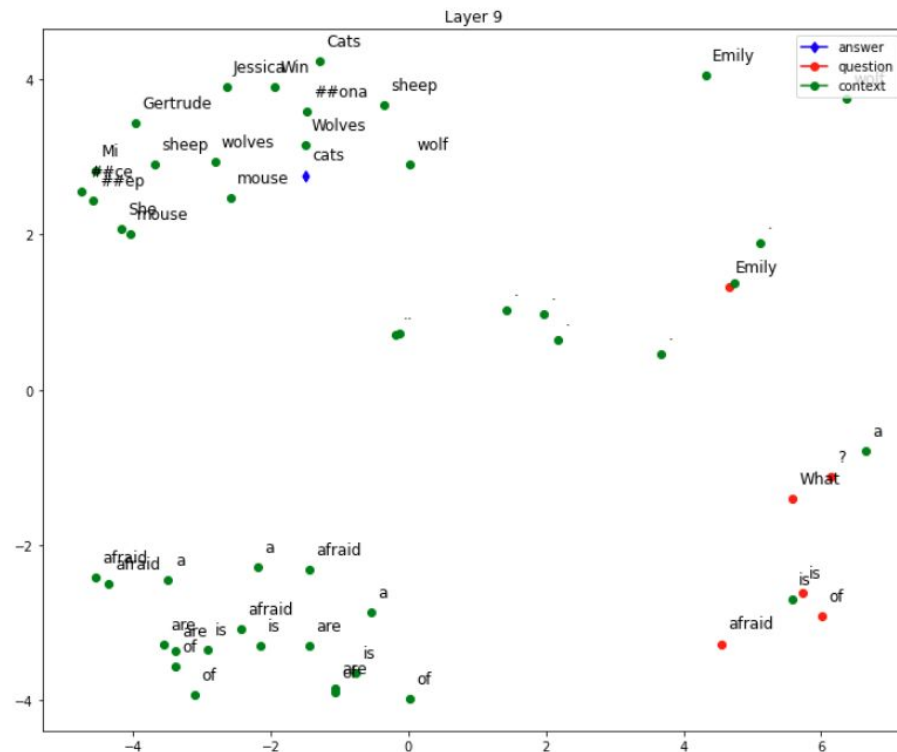
Question 25 to 27

Visualize embedding across layers of bert using PCA (Principal Component Analysis)

Objective:

- (1) How does bert solve question answering?
- (2) Change of embedding before and after fine-tuning

You only need to change code in the section "TODO"!



Embedding Analysis

Question 28 to 30

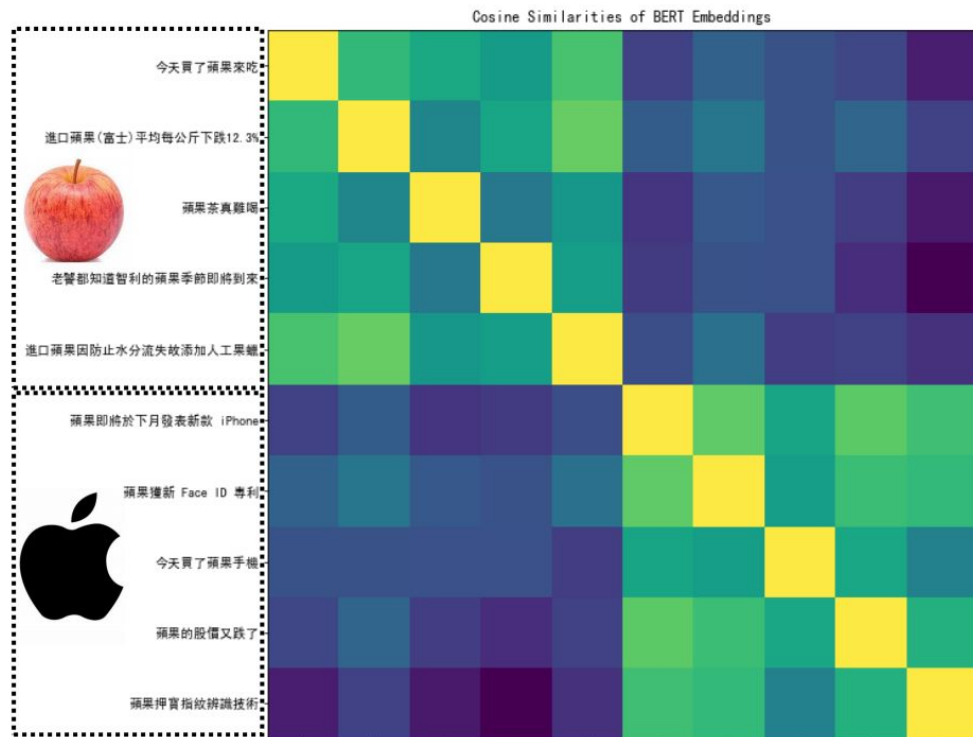
Compare output embedding of bert

using (1) Euclidean distance

(2) Cosine similarity

Objective:

- (1) Observe different meanings for the same word
- (2) Observe representation in different layers



You only need to change code in the section "TODO"!

Grading

- 30 multiple choice questions
- CNN: 20 questions
 - 0.3 pt for each question
- BERT: 10 questions
 - 0.4 pt for each question
- You have to choose ALL the correct answers for each question

Submission

- No late submission!
- Deadline: 2021/5/28 23:59

Reminder

- Please don't change the original code, unless the question request you to do so.
- If there is any confusion, email the TA with the subject "[HW9] ..."

Links

- Code:

[\[Colab\]](#)

- Questions:

[\[INTU COOL\]](#)

If any questions, you can ask us via...

- NTU COOL (recommended)
 - <https://cool.ntu.edu.tw/courses/4793>
- Email
 - ntu-ml-2021spring-ta@googlegroups.com
 - The title **must** begin with “[hw9]”
- TA hours
 - Each Monday 19:00~21:00 @Room 101, EE2 (電機二館101)
 - Each Friday 13:30~14:20 Before Class @Lecture Hall (綜合大講堂)
 - Each Friday During Class

李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

中英双语字幕



课件

一键打包下载



笔记

官方笔记翻译



代码

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1fM4y137M4>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/ntu-hylee-ml>

机器学习
深度学习

批次标准化

Auto-encoder

卷积神经网络

生成式对抗网络

GAN

自监督

强化学习

元学习

学习率

自注意力机制

Transformer

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



微信公众号

资料下载方式 2: 扫码点击**底部菜单栏**
称为 **AI 内容创作者**? 回复 [添砖加瓦]