HW9

Graded

Student
黃資芸

Total Points
10 / 10 pts

Question 1
Lime — **0.3** / 0.3 pts

+ 0 pts Incorrect

✔ + 0.3 pts Correct

Question 2
Lime — **0.3** / 0.3 pts

+ 0 pts Incorrect

✔ + 0.3 pts Correct

Question 3
Lime — **0.3** / 0.3 pts

✔ + 0.3 pts Correct

+ 0 pts Incorrect

Question 4
Lime — **0.3** / 0.3 pts

✔ + 0.3 pts Correct

+ 0 pts Incorrect

Question 5
Saliency Map — **0.3** / 0.3 pts

✔ + 0.3 pts Correct

+ 0 pts Incorrect

Question 6
Saliency Map — **0.3** / 0.3 pts

✔ + 0.3 pts Correct

+ 0 pts Incorrect

Question 7

Saliency Map **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 8

Saliency Map **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 9

Saliency Map **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 10

Smooth Grad **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 11

Smooth Grad **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 12

Smooth Grad **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 13

Smooth Grad **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 14

Filter Explanation **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 15

Filter Explanation **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 16

Fiter Explanation **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 17

Filter Explanation **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 18

Integrated Gradient **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 19

Integrated Gradient **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 20

Integrated Gradient **0.3** / 0.3 pts

✔ **+ 0.3 pts** Correct

+ 0 pts Incorrect

Question 21

Attention Visualization **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Question 22

Attention Visualization **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Question 23

Attention Visualization **0.4** / 0.4 pts

23.1 Attention Visualization **0.2** / 0.2 pts

✔ **+ 0.2 pts** Correct

+ 0 pts Incorrect

23.2 Attention Visualization **0.2** / 0.2 pts

✔ **+ 0.2 pts** Correct

+ 0 pts Incorrect

Question 24

Attention Visualization **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Question 25

Embedding Visualization **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Question 26

Embedding Visualization **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Question 27

Embedding Visualization **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Question 28

Embedding Analysis **0.4** / 0.4 pts

28.1 **Embedding Analysis** **0.2** / 0.2 pts

✔ **+ 0.2 pts** Correct

+ 0 pts Incorrect

28.2 **Embedding Analysis** **0.2** / 0.2 pts

✔ **+ 0.2 pts** Correct

+ 0 pts Incorrect

Question 29

Embedding Analysis **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Question 30

Embedding Analysis **0.4** / 0.4 pts

✔ **+ 0.4 pts** Correct

+ 0 pts Incorrect

Q1 Lime
0.3 Points

請觀察圖片編號2，包含了中央的牛奶瓶與左下角的草莓。在使用 Lime 套件之後的結果中，請問圖片中的這兩個部分怎麼影響 model 做出分類？
Please observe picture number 2, which contains the milk bottle in the center and the strawberry in the lower-left corner. In the results after using the Lime package, how do these two parts of the picture affect the classification of the model?

- ○ 中央的牛奶瓶為主要的正相關 / The milk bottle in the center is the main positive correlation
- ◉ 左下角的草莓為主要的負相關 / The strawberry in the lower-left corner is the main negative correlation
- ○ 中央的牛奶瓶為主要的負相關 / The milk bottle in the center is the main negative correlation
- ○ 左下角的草莓為主要的正相關 / The strawberry in the lower-left corner is the main positive correlation

Q2 Lime
0.3 Points

請問 Lime 套件的顏色代表什麼意義？（選兩個選項）
What does the color of the Lime package mean? (Choose two options)

- ☑ 綠色代表正相關 / Green means the positive correlation
- ☑ 紅色代表負相關 / Red represents the negative correlation
- ☐ 綠色代表負相關 / Green represents the negative correlation
- ☐ 紅色代表正相關 / Red means the positive correlation

Q3 Lime
0.3 Points

請問 Lime 套件是如何找出圖片各個部分對於物件辨識模型的相關性？
How does the Lime package discover the correlation between each part of the picture with the model's judgment?

- ○ 加入 noise 觀察輸出結果的改變 / Adding noise to observe the difference in the outputs of the model
- ○ 根據Loss對圖片的偏微分值 / According to the partial differential value of the loss to the picture
- ○ 根據圖片中各個部分的顏色 / According to the colors in each part of the picture
- ◉ 將圖片切成小塊觀察模型判斷的變化 / Slice the pictures into small components to observe the difference in the model's judgment

Q4 Lime
0.3 Points

請觀察圖片編號 3 使用 Lime 套件的結果，請問何者正確？
Please observe the result of using the Lime package for picture number 3. Which option is correct?

- ◉ 有一部分蛋糕對於 model 來說是正相關，有些則是負相關 / Some part of the cake is positively correlated to the model, and some part is negatively correlated
- ○ 蛋糕對於 model 來說完全是正相關的依據 / The cake is completely positively related to the model
- ○ 蛋糕不是 model 的判斷依據 / The cake is not the basis for the judgment of the model
- ○ 蛋糕對於 model 來說完全是負相關的依據 / The cake is completely negatively related to the model

Q5 Saliency Map
0.3 Points

請問在畫 Saliency Map 時，為何要對每張圖片的 gradient 做 normalization?
When drawing the Saliency Map, why do we need to normalize each image's gradient?

- ○ 增加不同圖片之間 gradient scale 不同的影響 / Increase the effect of the gradient scale between different pictures
- ◉ 降低不同圖片之間 gradient scale 不同的影響 / Reduce the effect of the gradient scale between different pictures
- ○ 有沒有做都沒有影響 / It doesn't matter whether we do it or not
- ○ 讓 Saliency 分布更為顯著 / Make saliency distribution more significant
- ○ 讓 Saliency 分布更為均勻 / Make saliency distribution more even

Q6 Saliency Map
0.3 Points

請問圖片編號 8 的 Saliency map 中，紅點分布的情況最接近下列何者？
In the Saliency map of picture number 8, which of the following is closest to the distribution of red dots?

- ○ 明顯分布於圖片中筷子的位置 / Obviously distributed in the position of the chopsticks in the picture
- ○ 明顯分布於圖片中盤子的位置 / Obviously distributed in the position of the plate in the picture
- ◉ 明顯分布於圖片中生魚片的位置 / Obviously distributed in the position of the sashimi in the picture
- ○ 明顯分布於圖片中桌子的位置 / Obviously distributed in the position of the table in the picture

Q7 Saliency Map
0.3 Points

請問 Saliency Map 是將下列何者具象化？
Which of the following does Saliency Map visualize?

- ○ Input tensor 對 loss 的偏微分值 / The partial differential value of input tensor to loss
- ○ Model parameter 對 loss 的偏微分值 / The partial differential value of model parameter to loss
- ○ Loss 對 model parameter 的偏微分值 / The partial differential value of loss to model parameter
- ◉ Loss 對 input tensor 的偏微分值 / The partial differential value of loss to input tensor

Q8 Saliency Map
0.3 Points

請問圖片編號 1 的 Saliency Map 中，紅點的分佈最像下列哪一種？
In the Saliency map of picture number 1, which of the following is closest to the distribution of red dots?

- ○ 角錐體 Pyramid
- ○ 四面體 Tetrahedron
- ○ 圓柱體 Cylinder
- ◉ 立方體 Cube
- ○ 圓球體 Round sphere

Q9 Saliency Map
0.3 Points

請問哪一張圖片使用 Lime 套件的結果正相關分布與 Saliency Map 最明顯的不同？
Which picture has the most significant difference between the positive correlation distribution using the Lime package and Saliency Map?

- ○ 圖片編號 7 picture number 7
- ○ 圖片編號 8 picture number 8
- ◉ 圖片編號 5 picture number 5
- ○ 圖片編號 6 picture number 6
- ○ 圖片編號 4 picture number 4

Q10 Smooth Grad
0.3 Points

請問 Smooth Grad 是藉由甚麼方式解釋 model 的判斷？
How does Smooth Grad explain the model's judgment?

- ○ 平均多張不同圖片的結果以觀察 model 產生的 Saliency Map / Average the results of multiple different pictures to observe the Saliency Map generated by the model
- ◉ 隨機加入 noise 觀察 model 產生的 Saliency Map / Randomly add noise to observe the Saliency Map generated by the model
- ○ 隨機加入 noise 觀察 model 的輸出結果改變 / Randomly add noise to observe the changes in the output of the model
- ○ 平均多張不同圖片的結果以觀察 model 的輸出 / Average the results of multiple different pictures to observe the output of the model

Q11 Smooth Grad
0.3 Points

請比較 Saliency Map 與 Smooth Grad 產生的結果，下列何者正確？
Please compare the results produced by the Saliency Map and Smooth Grad. Which of the following is correct?

- ○ 圖片編號 4 的 Saliency Map 和 Smooth Grad 強調的位置不同 / The Saliency Map of the picture number 4 is different from the prominent position of Smooth Grad
- ○ 整體來說，Saliency Map 強調的部分更能清楚地呈現圖片中食物的位置 / Overall, the highlighted part of Saliency Map can present the position of the food in the picture more clearly
- ○ 圖片編號 2 的 Saliency Map 和 Smooth Grad 強調的位置相同 / The Saliency Map of the picture number 2 is the same as the prominent position of Smooth Grad
- ◉ 整體來說，Smooth Grad 強調的部分更能清楚地呈現圖片中食物的位置 / Overall, the highlighted part of Smooth Grad can present the position of the food in the picture more clearly

Q12 Smooth Grad
0.3 Points

請問在 Smooth Grad 計算完成後，沒有使用 normalization 會造成什麼結果？
After the calculation of Smooth Grad, what will happen if normalization is not used?

- ○ 食物的輪廓更加清楚 / The outline of the food is clearer
- ○ 與有使用 normalization 的結果一樣 / The same as the result of using normalization
- ◉ 無法觀察到亮點部位 / Cannot observe bright spots
- ○ 亮點變暗，暗點變亮 / Bright spots darken, dark spots brighten
- ○ 產生亮點的區域不同 / The areas where the bright spots are generated are different

Q13 Smooth Grad
0.3 Points

請觀察圖片編號 7 的 Smooth Grad 結果，下列何者正確？
Please observe the Smooth Grad result of picture number 7. Which of the following is correct?

- ○ 亮點主要為米腸 / The highlight is mainly the rice sausage
- ◉ 亮點主要為麵 / The highlight is mainly the noodle
- ○ 亮點主要為青菜 / The highlight is mainly the vegetable
- ○ 亮點主要為盤子 / The highlight is mainly the plate

Q14 Filter Explanation
0.3 Points

請觀察圖片編號 2 在 cnnid=15, filterid=0 的 filter activation 結果，請問圖片中的哪部分最**不能** activate 這個 filter？
Please observe the result of filter activation with picture number 2 at cnnid=15, and filterid=0. Which part of the picture **can not** activate the filter most?

- ○ 牛奶瓶輪廓 / the contour of the milk bottle
- ◉ 背景 / background
- ○ 陰影 / shadow
- ○ 草莓 / strawberry

Q15 Filter Explanation
0.3 Points

請觀察較靠近輸出端與較靠近輸入端的 CNN layer 的 filter activation 有什麼差異？(選 2 個選項)
Please observe the difference between the filter activation of the CNN layer closer to the output and that closer to the input? (Choose two options)

- [ ] 兩者 activate 的位置幾乎不同 / The position of activation is almost different
- [ ] 靠近輸出端的比較清晰 / The filter activation closer to the output is clearer
- [x] 兩者 activate 的位置大致相似 / The position of activation is roughly similar
- [x] 靠近輸出端的比較模糊 / The filter activation closer to the output is vaguer

Q16 Fiter Explanation
0.3 Points

Filter Explanation 中我們觀察了哪些情況？(選 2 個選項)
What have we observed in Filter Explanation? (Choose two options)

- [x] 什麼樣的圖片最容易 activate 特定的 filter / What kind of picture is the easiest to activate a specific filter
- [ ] 哪一層 filter 對判斷食物種類最有幫助 / Which level of filter is most helpful for judging the type of food
- [x] 圖片的哪些位置會 activate 特定的 filter / Which parts of the picture activate a specific filter
- [ ] 哪一層 filter 最容易被特定的圖片 activate / Which layer of the filter is most likely to be activated by a specific picture

Q17 Filter Explanation
0.3 Points

請問我們觀察 cnnid=21 的 filter visualization，是觀察 model 的哪一種 layer 的輸出？
When we observe the filter visualization with cnnid=21, what kind of layer output of the model are we observing?

- ○ Conv2d()
- ◉ BatchNorm2d()
- ○ Linear()
- ○ ReLU()
- ○ MaxPool2d()

Q18 Integrated Gradient
0.3 Points

請問在使用 Integrated Gradient 方法時把 generate_images_on_linear_path 的 steps 調得更大，結果會產生什麼差異？
What difference in the results will be made if the steps of generate_images_on_linear_path are tuned larger when using the Integrated Gradient method？

- ○ 亮點變暗，暗點變亮 / Bright spots darken, dark spots brighten
- ○ 觀察不到亮點 / No bright spots are observed
- ○ 亮點隨機出現 / Bright spots appear randomly
- ◉ 沒有顯著差異 / No significant difference

Q19 Integrated Gradient
0.3 Points

請問 Integrated Gradient 中 generate_images_on_linear_path 這個函數的功能為何？
What is the functionality of generate_images_on_linear_path in Integrated Gradient?

- (●) 在原圖片與 baseline 之間產生連續的 samples / Generate continuous samples between the original picture and the baseline
- ( ) 根據原圖片產生新的一群隨機圖片 / Generate a new group of random pictures based on the original picture
- ( ) 把原圖片線性平移產生一張新圖片 / Linearly translate the original picture to generate a new picture
- ( ) 根據原圖片產生一張 baseline 的圖片 / Generate a baseline picture based on the original picture

Q20 Integrated Gradient
0.3 Points

請觀察圖片編號 3 Integrated Gradient 的結果，請問 model 判斷的依據可能是哪些 pixels？
Please observe the result of picture number 3 in Integrated Gradient. What pixels may be the basis for model judgment?

- ( ) 叉子 / fork
- (●) 蛋糕 / cake
- ( ) 盤子 / plate
- ( ) 桌子 / table

Q21 Attention Visualization

0.4 Points

請使用網站 https://exbert.net/exBERT.html 及模型 [bert-based-cased] 分析以下句子：
The police officer is popular with the residents because she is very generous and kind.
對於模型中12層，每層中的第12個attention head (i.e. layer 1 head 12, layer 2 head 12, layer 3 head 12, ..., layer 12 head 12)，下列哪一個功能可能存在？(可能超過一個答案)

Please use the website https://exbert.net/exBERT.html and the model "bert-based-cased" to visualize the sentence "The police officer is popular with the residents because she is very generous and kind."
For attention head 12 across all 12 layers in the model (i.e. layer 1 head 12, layer 2 head 12, layer 3 head 12, ..., layer 12 head 12), which of the following functionalities is most likely to exist? (maybe more than one answer)

- [x] Attend to the previous token
- [x] Attend to the period (句號)
- [x] Attend to special token
- [x] Attend to the same token
- [x] Attend to the next token

Q22 Attention Visualization
0.4 Points

請使用網站 https://exbert.net/exBERT.html 及模型 [bert-based-cased] 分析以下句子：
The police officer is popular with the residents because she is very generous and kind.
在 [officer] 一字被蓋掉，換成 [MASK] token 的情況下，搜尋 [Wizard of Oz] 中與 [MASK] 最相似的embedding。請問模型從哪一層開始成功預測被蓋掉的字的詞性？(成功預測的定義是50個字中最常出現的詞性與被蓋掉的字詞性相同)

Please use the website https://exbert.net/exBERT.html and the model “bert-based-cased” to visualize the sentence “The police officer is popular with the residents because she is very generous and kind.”
Mask the word “officer” and search in the corpus “Wizard of Oz” for embeddings most similar to the masked token “[MASK]”. After which layer does the model successfully predict part of speech of the masked word? (the definition of predict successfully is that the most frequent part of search among the 50 words is the same as the masked word)

- ○ Layer 7
- ○ Layer 11
- ○ Layer 9
- ○ Layer 10
- ○ Layer 8
- ◉ Layer 12

Q23 Attention Visualization
0.4 Points

請使用網站 https://exbert.net/exBERT.html 及模型 [bert-based-cased] 分析以下句子：
The police officer is popular with the residents because she is very generous and kind.

Please use the website https://exbert.net/exBERT.html and the model “bert-based-cased” to visualize the sentence “The police officer is popular with the residents because she is very generous and kind.”

Q23.1 Attention Visualization
0.2 Points

當句子中的 [she] 被蓋掉，模型在蓋掉的地方預測哪一個字？
Which token does the model predict in the masked position if “she” in the sentence is masked?

- (●) he
- ( ) she
- ( ) it

Q23.2 Attention Visualization
0.2 Points

在句子中的 [she] 被蓋掉的情況下，如果把句子中的 [officer] 也蓋掉，在 [she] 的位置預測 [he] 的機率減少了多少？
By how much does the probability of predicting “he” in the position of “she” decrease when “officer” is also masked? (“she” is masked during the comparison)

- (●) 0.81
- ( ) 0.18
- ( ) 0.99
- ( ) 0.01

Q24 Attention Visualization
0.4 Points

請使用網站 https://exbert.net/exBERT.html 及模型 [bert-based-cased] 分析以下句子：
The police officer is popular with the residents because she is very generous and kind.
下列哪一個attention head可能有指代消解的功能？
(指代消解的例子：湯姆喜歡珍妮，因為她很可愛。她指的是珍妮)

Please use the website https://exbert.net/exBERT.html and the model “bert-based-cased” to visualize the sentence “The police officer is popular with the residents because she is very generous and kind.”
Which of the following attention head may have the functionality of coreference resolution?
(example of coreference resolution: Tom likes Jennie because she is cute. She refers to Jennie)

- ○ Layer 7 head 10
- ◉ Layer 5 head 10
- ○ Layer 4 head 10
- ○ Layer 6 head 10
- ○ Layer 8 head 10

Q25 Embedding Visualization
0.4 Points

在模型的哪一層中，答案的 embedding 與其他 embedding 離最遠？
In which layer of the model, the embedding of the answer is the furthest away from the other embeddings?

- ◉ Layer 11
- ○ Layer 2
- ○ Layer 8
- ○ Layer 5

Q26 Embedding Visualization
0.4 Points

模型中的哪幾層可能在負責[在文章中尋找與問題有關的資訊]？
Which layers of the model may perform the step “Matching questions with relevant information in context”?

- ○ Layer 1 to 3
- ○ Layer 10 to 12
- ○ Layer 4 to 6
- ◉ Layer 7 to 9

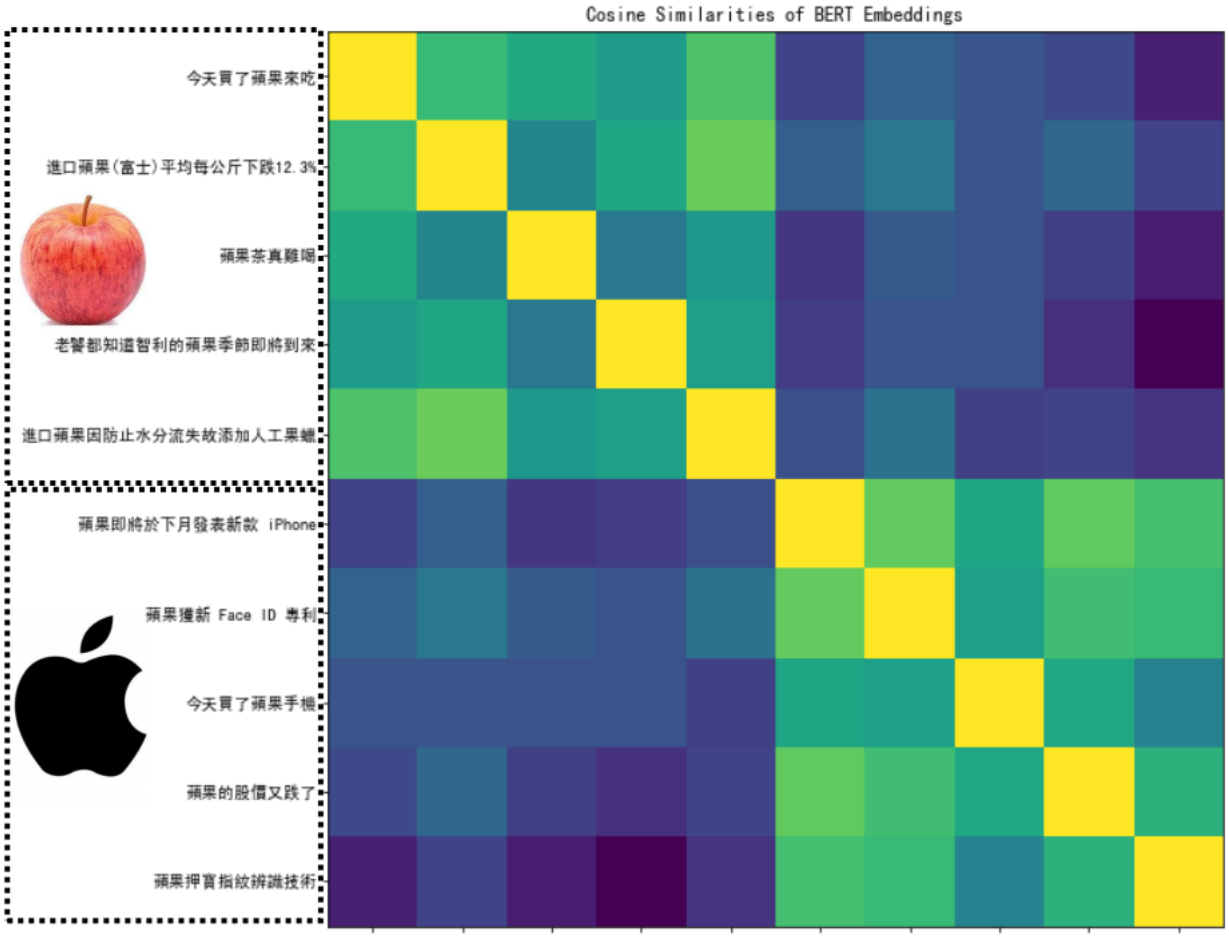
Q27 Embedding Visualization
0.4 Points

模型中的哪幾層可能在負責[將類似的文字分群(根據文字在文章中的關係)]？
Which layers of the model may perform the step “Clustering similar words together (based on the relation of words in context)?

- ○ Layer 1 to 3
- ○ Layer 7 to 9
- ◉ Layer 4 to 6
- ○ Layer 10 to 12

Q28 Embedding Analysis
0.4 Points

請找出作業投影片中的圖片(也是老師上課投影片的圖片)是如何生成的。
Please reproduce the picture in the homework slide (which is also the picture in the professor's slide)



Q28.1 Embedding Analysis
0.2 Points

圖片使用的比較指標是？
Which metric is used for comparison?

- ○ 歐氏距離 / Euclidean distance
- ◉ 餘弦相似度 / Cosine similarity

Q28.2 Embedding Analysis
0.2 Points

圖片在比較哪一個字/詞的 Embedding? (註：正確答案的圖片顏色會幾乎相同)
Which word/phrase is used for comparison? (For the correct answer, the colors of the image are nearly the same)

- ○ 蘋果的embedding / Embedding of 蘋果
- ◉ 果的embedding / Embedding of 果 only
- ○ 蘋的embedding / Embedding of 蘋 only

Q29 Embedding Analysis
0.4 Points

使用餘弦相似度或歐氏距離作為比較的指標。請比較句子 [今天買了蘋果手機] 和 [蘋果獲得新Face ID專利] 中，兩個[果]字的相似度。請問 Embedding 從第 1 層到第 11 層(請忽略最後一層)，相似度的趨勢為何？

Use the metric “Cosine similarity” or “Euclidean distance” for comparison. Compare the word embedding of [果] in the sentence [今天買了蘋果手機] and that in the sentence [蘋果獲得新Face ID專利]. What is the trend of similarity from embedding output from layer 1 to layer 11? (please exclude the last layer)

- ○ 下跌 / Decrease
- ○ 上升 / Increase
- ◉ 先跌後升 / Decrease, then increase
- ○ 先升後跌 / Increase, then decrease

Q30 Embedding Analysis
0.4 Points

使用餘弦相似度或歐氏距離做比較的指標。請問在第 0 層(input embedding)，不同句子中的[果]一字之間的相似度，與下列何者有關？(可能有超過一個答案)
Which of the following(s) affects the similarity of [果] between different sentences at the embedding output in layer 0? (maybe more than one answer)

- [x] 果在句子中的位置position of 果 in sentences
- [ ] 果在句子中的意思contextual meaning of 果 in sentences
- [ ] 果所在句子的長度 length of sentences
- [ ] 果在句子中出現的次數occurrence of 果 in sentences