Week 8: Assignment

1) Which of the following best describes the purpose of pixel attribution methods in image classification by neural networks?
    a) To increase the resolution of an image by modifying pixel values.
    b) To highlight the pixels that were most relevant for the neural network's decision in classifying an image.
    c) To reduce the noise in an image by adjusting irrelevant pixels.
    d) To segment the image into different regions based on pixel similarity.

2) Which of the following is NOT a name commonly associated with pixel attribution methods?
    a) Saliency map
    b) Sensitivity map
    c) Feature attribution
    d) Convolution map

3) Which of the following statements is true regarding pixel attribution methods in image classification?
    a) SHAP and LIME are gradient-based methods that compute the gradient of the prediction with respect to input features.
    b) Gradient-based methods generate explanations by manipulating parts of the image to see how it affects the classification.
    c) Occlusion-based methods manipulate parts of the image, such as blocking or altering pixels, to understand their influence on the model's decision.
    d) All pixel attribution methods require model-specific adjustments to function correctly.

4) Which of the following is a key difference between StyleGAN2 and StyleGAN3?
    a. StyleGAN2 is fully equivariant to translation and rotation,improving the identification of important properties.
    b. StyleGAN3 focuses on improving the attachment of details to the image surface, whereas StyleGAN2 struggles with internal representations.
    c. StyleGAN2 is better at identifying important properties due to its fully equivariant nature.
    d. StyleGAN3 is fully equivariant to translation and rotation, improving the identification of important properties.

5) What is the primary purpose of adding noise to the image in the Smooth Grad method?
    a) To enhance the resolution of the image.
    b) To create multiple variations for averaging pixel attribution maps.
    c) To reduce the effect of irrelevant classes.
    d) To increase the complexity of the gradient computation.

6) How does Guided BackProp differ from standard backpropagation in generating saliency maps?
   a) It only considers positive gradients by zeroing out negative activations and gradients.
   b) It back propagates gradients with all activations zeroed out.
   c) It focuses on highlighting both negative and positive contributions.
   d) It requires padding 1 to the image before backpropagation.

7) What does a lack of change in saliency maps after randomizing the layers indicate?
   a) The saliency maps are highly accurate in reflecting the model's learning.
   b) The saliency maps cannot be deceptive.
   c) The saliency maps are unreliable and may not accurately capture the model’s learned features.
   d) The saliency maps provide detailed visualizations of the model's internal mechanisms

8) What is a key feature of LIME (Local Interpretable Model-agnostic Explanations)?
   a) It requires access to the internal workings of the model to generate explanations.
   b) It only works with tabular data and cannot be applied to text.
   c) It provides explanations that are globally faithful across all predictions.
   d) It can be used with any black box model, regardless of the model's internal structure.

9) What is the primary basis of SHAP (SHapley Additive exPlanations) for generating explanations?
   a) It employs a game theoretic approach to allocate credit and explain predictions.
   b) It uses a neural network to generate explanations based on model weights.
   c) It applies statistical sampling methods to estimate the importance of features.
   d) It utilizes clustering techniques to group similar data points for explanation.

10) How do ProtoPNet models determine which patches are most important for classification?
   a) By evaluating the overall texture patterns of images.
   b) By using statistical correlation between different patches of images.
   c) By identifying and using patches that are representative or prototypical of each class.
   d) By performing dimensionality reduction on the image data to find key features.

11) Why is probing important even when a model shows strong performance on a task?

a) To check if the model is using irrelevant data for making predictions.
b) To verify if the model's high accuracy is due to performing specific subtasks effectively.
c) To understand whether the model is overfitting to the training data.
d) To determine the computational efficiency of the model during training and inference.

12) Which of the following best describes the TokFSM dataset?

a) It is a dataset focused on image classification.
b) It is a dataset for natural language generation.
c) It is a dataset for reinforcement learning tasks.
d) It is an algorithmic sequence modeling dataset.

13) How do we identify "pure" codes in a codebook model?

a) By checking if they activate on only one bigram or trigram
b) By evaluating their impact on training time
c) By measuring their effect on model accuracy
d) By analyzing their computational complexity

14) Which of the following methods belong to the occlusion- or perturbation-based category of pixel attribution methods?

a) Gradient Class Activation Mapping (Grad-CAM)
b) Integrated Gradient
c) DeepLIFT
d) SHAP
e) LIME