DataLab Preparation (Week 7, DataLab I, Wednesday)

2.1b Which step is NOT part of generating a vanilla gradient visualization for a deep learning model?

Performing a forward pass to obtain the model's prediction for an input image.

[x]Randomly perturbing pixels in the input image to observe changes in output.

Calculating the gradient of the model's output with respect to the input image pixels.

Using backpropagation to compute the gradient values.

2.1c What does it mean when a pixel has a very high gradient in the context of vanilla gradients?

The pixel is highly compressed and needs to be decompressed for accurate model predictions.

The pixel's color needs to be adjusted for the model to recognize the input correctly.

[x]Small changes in the pixel's value could lead to significant changes in the model's prediction score.

The pixel is irrelevant to the model's prediction and can be ignored in further analysis.

2.1d What does a spread-out gradient pattern in a vanilla gradient image tell us about how the model makes its choices?

[x]The model's decision-making process may be influenced by a wide range of features across the input.

The model is focusing on specific, localized features in the input.

The model is highly confident in its predictions.

The model is likely underfitting to the training data.

2.2b What does saturation mean in the context of integrated gradients, and how does it affect the interpretation of model predictions?

Saturation refers to the model's ability to process high-resolution images, increasing accuracy.

[x]It describes a situation where changing an input feature significantly does not change the model's output, indicating that the model's prediction is insensitive to that feature.

Saturation is the process of increasing the brightness of input images to make model predictions more confident.

It means the model has reached its maximum storage capacity for input features, requiring dimensionality reduction techniques.

2.2c What is a baseline in the context of integrated gradients, and why is it important?

A high-performance model used to compare against the current model.

The initial weights of the neural network before training begins.

[x]A reference input used to measure the contribution of each feature to the model's prediction.

The maximum value of gradients obtained during model training.