Quiz for Audio Features Tutorial

Hint: For each question, one or more options can be correct.

1. What is the main purpose of using Mel-Frequency Cepstral Coefficients (MFCCs) in audio processing?
    a) To compress the audio file size
    b) To extract features related to the human perception of sound
    c) To filter out noise from the audio
    d) To convert audio into a visualizable format
2. Why might one have to pad audio files before using them in machine learning models?
    a) To make all audio clips the same length
    b) To improve the audio quality
    c) To synchronize audio with video
    d) To increase the computational efficiency
3. What additional audio features could, in principle, be helpful for emotion classification for audio (apart from MFCCs)?
    a) Spectral contrast and spectral centroid
    b) Pitch and tempo
    c) Beat and rhythm
    d) Volume
4. Why does one (sometimes) remove the mean and scale the audio features to unit variance before the machine learning process?
    a) To ensure that all features contribute equally to the model
    b) To prevent features with larger scales from dominating the learning process
    c) To reduce the likelihood of overfitting in models sensitive to feature scales
    d) To comply with the assumptions of certain algorithms that expect normalized data (side note: example?)
5. Using the raw audio waves with a Support Vector Machine is
    a) Will necessarily improve the accuracy relative to features such as MFCCs
    b) Can be challenging due to the high dimensionality
    c) Computationally more efficient than using MFCCs
    d) Provides a high interpretability to the model decisions
6. What effect did downsampling the audio have on the length of the raw waves?
    a) It increased the maximum length of the waves
    b) It decreased the maximum length of the waves

c) It standardized the length of all waves

d) It had no effect on the length of the waves

7. The RAVDESS dataset of emotional speech provides recordings from different actors

a) To capture a wide range of emotional expressions and nuances

b) Because recordings from multiple actors reduce background noise

c) To ensure diversity in voice types, accents, and speech patterns

d) To increase the dataset size for better machine learning model training

8. What is the purpose of using a spectral centroid in audio analysis?

a) To determine the loudness of the audio

b) To identify the dominant frequency in the audio

c) To measure the "brightness" of the sound

d) To locate the position of the sound source

9. How does the use of dropout layers in a neural network model affect its performance?

a) Increases the model accuracy significantly

b) Reduces the risk of overfitting

c) Decreases the training time

d) Improves the ability of the model to generalize to new data