Recitation 0.13
Data Preprocessing

Introduction to Deep Learning
11-785/685/485

Mona Aman, Mengchun Zhang

Agenda

1. Signal Preprocessing
2. Image Preprocessing

Why Data Preprocessing is Important

Data preprocessing is a vital step in the machine learning and deep learning pipeline. Raw data, whether it's audio, image, or any other format, often contains inconsistencies, noise, or irrelevant information that can hinder model performance. Effective preprocessing helps by:

- Transforming raw data into a **clean**, **standardized**, and analysis-ready format.
- **Enhancing feature quality**, making it easier for models to learn from meaningful patterns rather than irrelevant or misleading data.
- **Reducing the risk of overfitting** by introducing consistency and appropriate augmentation.
- **Improving the speed and stability** of model convergence during training.

Proper preprocessing ensures that your deep learning models receive high-quality, uniform inputs, ultimately leading to better accuracy, generalization, and robustness.

Introduction to Signal Preprocessing

Signal Preprocessing

When you hear someone speaking, your brain doesn't just process the raw sound waves. Instead, it picks out meaningful features like pitch, rhythm, tone, and phonetic information. Similarly, machine learning models also need to extract these meaningful features from raw audio to make sense of it and learn effectively.

- **Problem with Raw Audio:**
    - Raw audio is often **high-dimensional**. For example, a 2-second clip sampled at 16 kHz gives us **32,000 data points**.
    - This creates several challenges:
        - **Computational complexity**: It's expensive to process so much data.
        - **Noise sensitivity**: Raw audio contains irrelevant noise and variations.
        - **Lack of structure**: The waveform doesn't directly capture the linguistic or acoustic features we care about.

Feature Extraction - MFCCs

F**eature Extraction: MFCCs**

One of the most important preprocessing techniques for speech is Mel-Frequency Cepstral Coefficients (MFCCs).

Here's why MFCCs are so important for speech processing:

1. **Perceptual Relevance:**
   They are based on the mel scale, which reflects how humans perceive sound frequencies.

2. **Dimensionality Reduction:**
   Instead of working with thousands of raw audio samples, MFCCs provide a compact representation — usually just 12-13 coefficients per time frame.

3. **Robustness:**
   MFCCs are less sensitive to noise and variations in recording conditions, making them a great feature for speech recognition tasks.

Introduction to Image Preprocessing

What is Image Preprocessing?

Before training any machine learning model to understand images, we need to **prepare the data** properly. This is called **data preprocessing**, and it's crucial for successful computer vision tasks.

- **Why is it Important?**
    - Real-world images can be **messy**.
    - They may have **varying sizes**, **lighting conditions**, and **noise**.
    - Feeding raw images into a model can result in **poor performance**.

Key Preprocessing Steps

- **Resizing Images**:
  Ensure all images are the same **size** so they can be fed into a model.

- **Normalization**:
  Adjust pixel values to a **standard range** (e.g., 0-1 or -1 to 1) for stable training.

- **Augmentation**:
  Add variety to the dataset by artificially altering images (rotation, flipping, scaling) to make the model more robust.

- **Denoising**:
  Reduce noise and irrelevant details to enhance important features.

Data Augmentation for Robustness

Why Augment Data?

- **More Data**: Generate more training examples from the original dataset.
- **Prevent Overfitting**: Introduce diversity so the model doesn't memorize the training data.

Common Augmentation Techniques:

- **Rotation, Flipping**: Change image orientation.
- **Scaling, Cropping**: Alter size or focus area.
- **Color Jittering**: Randomly adjust brightness, contrast, or saturation.