Data Normalization

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

January 2021

UNIVERSITY OF CALGARY

Outline

- Learning Goals
- Data Normalization Strategies
- Summary

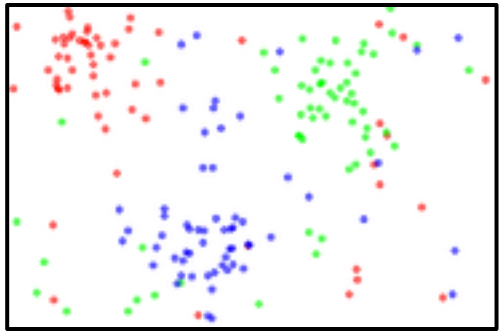
Learning Goals

- Learn the importance of data normalization
- Learn about the commonly used normalization strategies

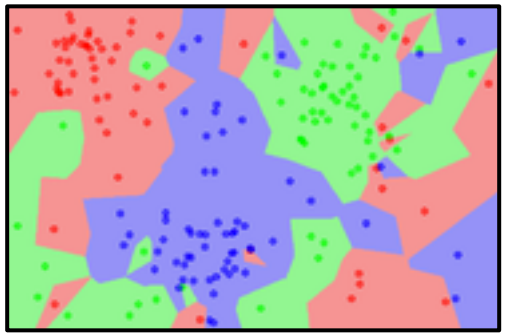
Data Normalization

- Reduce the influence of the different feature's scales (e.g., distance-based model where features have very different scales)
- Improves model training
- Need to be mindful of your data scale and your network output activation scale

Dataset



1-Nearest Neighbor



Notation

$$X = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1M} \\ x_{21} & x_{22} & \dots & x_{2M} \\ x_{31} & x_{32} & \dots & x_{3M} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NM} \end{bmatrix}}_{\text{N samples with M features}}$$

$$Y = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_5 \end{bmatrix}}_{\text{True Labels}}$$

$$\hat{Y} = \underbrace{\begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \widehat{y_3} \\ \dots \\ \widehat{y_N} \end{bmatrix}}_{\text{Predicted Labels}}$$

5

Notation

Notation	Meaning	Notation	Meaning	Notation	Meaning
X	Dataset	Y	Labels of dataset	$\widehat{Y}$	Predicted labels of dataset
X_{train}	Train set	Y_{train}	Labels of train set	$\widehat{Y_{train}}$	Predicted labels of train set
X_{val}	Validation set	Y_{val}	Labels of validation set	$\widehat{Y_{val}}$	Predicted labels of validation set
X_{test}	Test set	Y_{test}	Test set	$\widehat{Y_{test}}$	Predicted labels of test set

Notation

Notation	Meaning
X[i,:]	Sample i
X[:,j]	Feature j
X[i,j]	Feature j of sample i

Feature-wise Normalization

Min-max Normalization

$$X_{train}[:,i] = \frac{X_{train}[:,i] - \min(X_{train}[:,i])}{\max(X_{train}[:,i]) - \min(X_{train}[:,i])}$$

$$X_{val}[:,i] = \frac{X_{val}[:,i] - \min(X_{train}[:,i])}{\max(X_{train}[:,i]) - \min(X_{train}[:,i])}$$

$$X_{test}[:,i] = \frac{X_{test}[:,i] - \min(X_{train}[:,i])}{\max(X_{train}[:,i]) - \min(X_{train}[:,i])}$$

Standardization

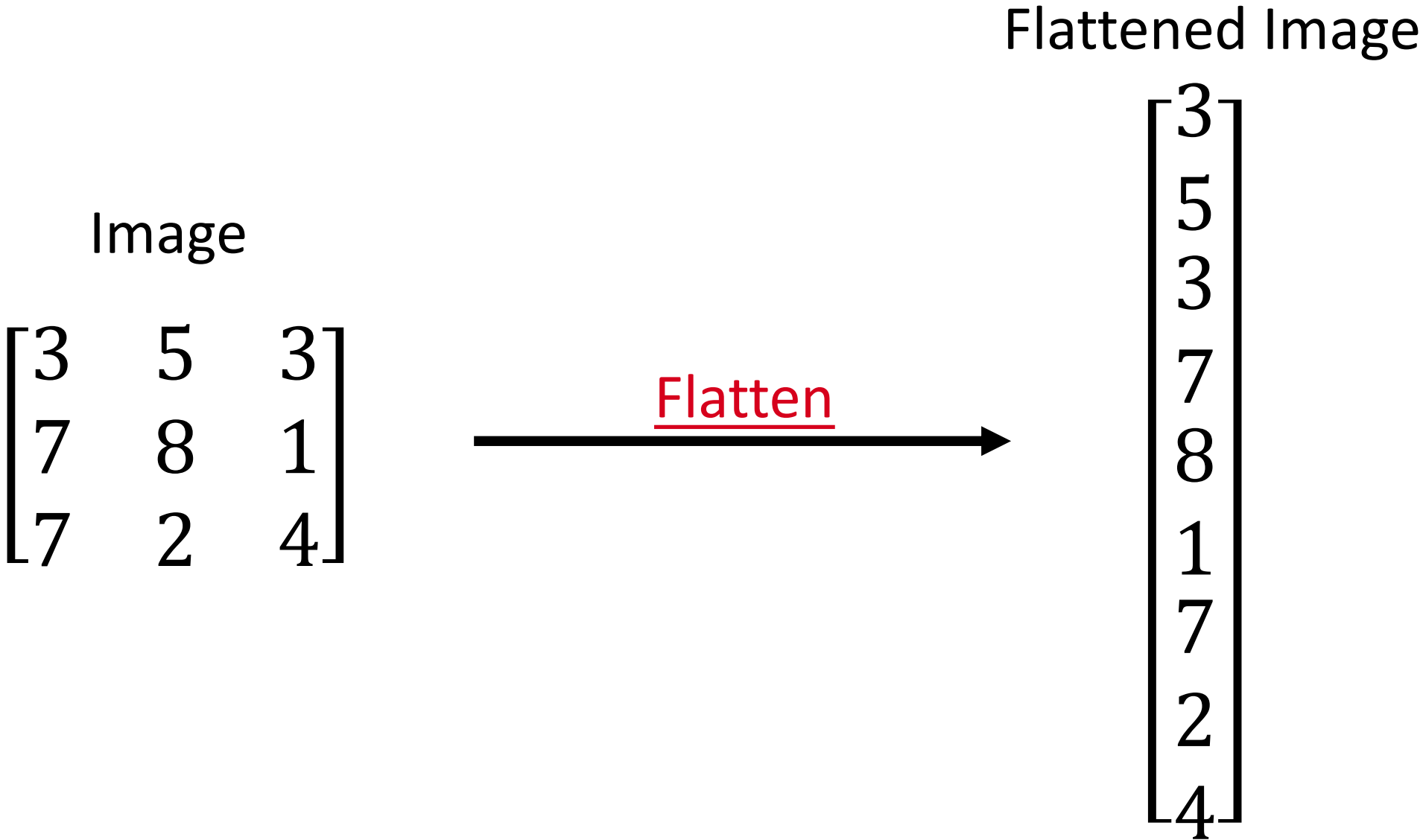
$$X_{train}[:, i] = \frac{X_{train}[:, i] - \mathrm{mean}(X_{train}[:, i])}{std(X_{train}[:, i])}$$

$$X_{val}[:, i] = \frac{X_{val}[:, i] - \mathrm{mean}(X_{train}[:, i])}{std(X_{train}[:, i])}$$

$$X_{test}[:, i] = \frac{X_{test}[:, i] - \mathrm{mean}(X_{train}[:, i])}{\mathrm{std}(X_{train}[:, i])}$$

Sample-wise Normalization

When working with locally correlated data…



Min-max Normalization (statistics of the training set)

$$X_{train} = \frac{X_{train} - \min(X_{train})}{\max(X_{train}) - \min(X_{train})}$$

$$X_{val} = \frac{X_{val} - \min(X_{train})}{\max(X_{train}) - \min(X_{train})}$$

$$X_{test} = \frac{X_{test} - \min(X_{train})}{\max(X_{train}) - \min(X_{train})}$$

Standardization (statistics of the training set)

$$X_{train} = \frac{X_{train} - \mathrm{mean}(X_{train})}{std(X_{train})}$$

$$X_{val} = \frac{X_{val} - \mathrm{mean}(X_{train})}{std(X_{train})}$$

$$X_{test} = \frac{X_{test} - \mathrm{mean}(X_{train})}{\mathrm{std}(X_{train})}$$

Sample-wise Normalization (statistics of the sample)

Min-max:

$$X[i,:] = \frac{X[i,:] - \min(X[i,:])}{\max(X[i,:]) - \min(X[i,:])}$$

Standardization:

$$X[i,:] = \frac{X[i,:] - mean(X[i,:])}{\text{std}(X[i,:])}$$

Other Normalization Strategies

- Batch Normalization
- Layer Normalization
- Output normalization

https://keras.io/api/layers/normalization_layers/

Summary

- Normalization is an essential step for properly training neural networks, special when you have features with different scales
- Three main types of normalization:
  - Feature-wise normalization based on the **statistics of the feature** in the training set
  - Sample-wise normalization based on the **statistics of all features** in the training set
  - Sample-wise normalization based on the **statistics of the sample**
- There is not one definite normalization method.

Thanks!