Analytics Vidhya

Tensor Parallelism

Instructor

Sourab Mangrulkar

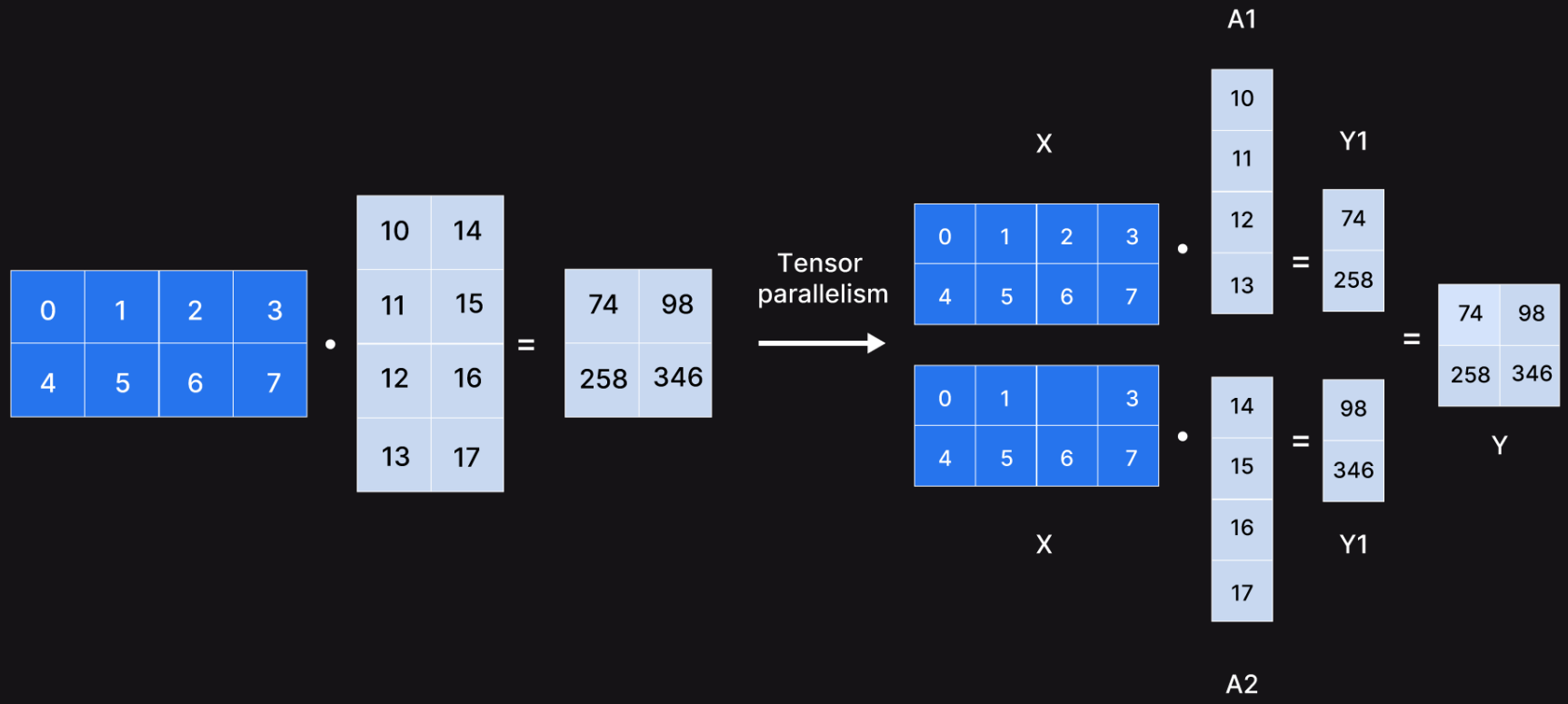
Machine Learning Engineer at

Creator of PEFT

Introduction

- Matrix Operations are the heart of neural networks
  - Dense Layers
  - Self attention layers

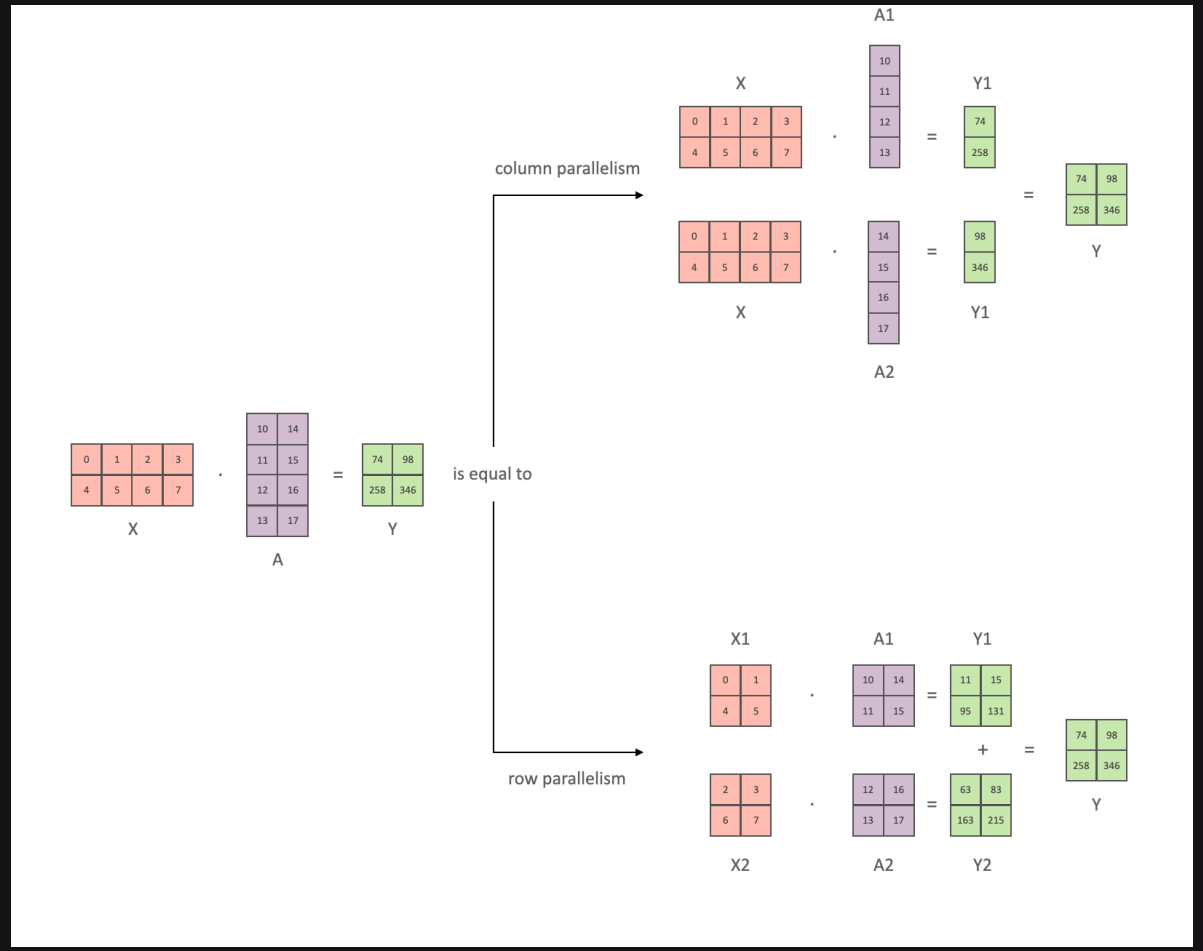
Tensor Parallelism



Analytics Vidhya

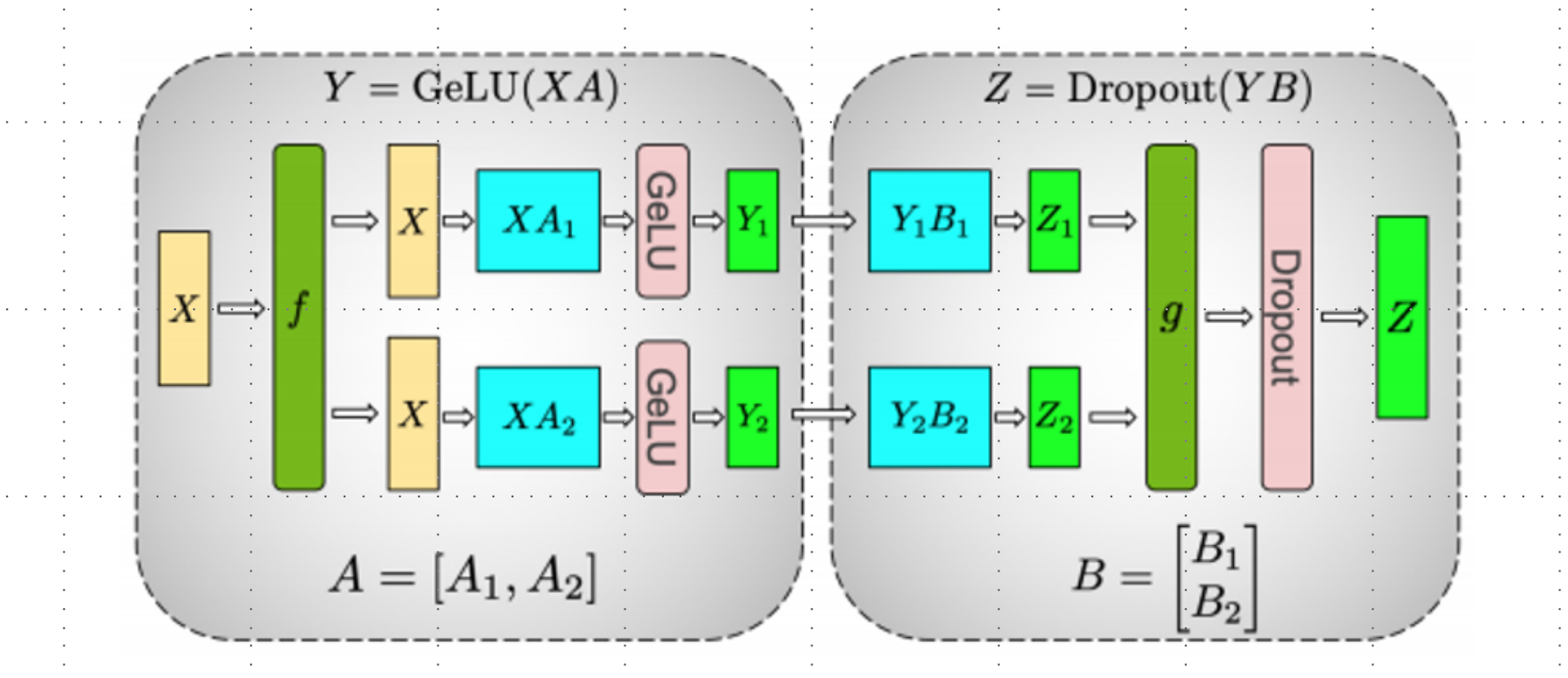
Tensor Parallelism

$$Y = GeLU(XA)$$

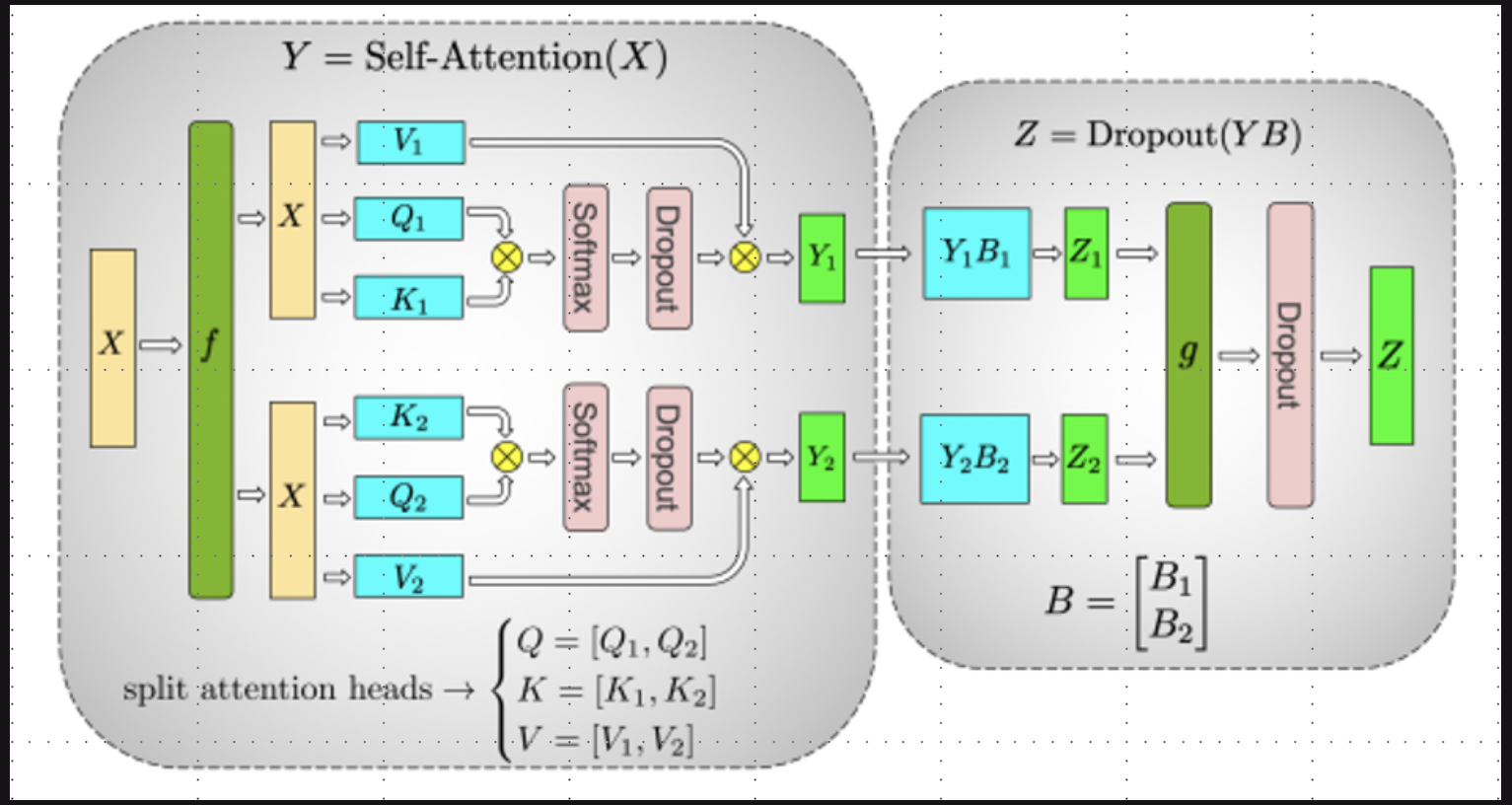


Analytics Vidhya

Tensor Parallelism



Tensor Parallelism



Thank You

Analytics Vidhya