Power Systems

IBM

Data Science 101

Chris Parsons

14 February
2018

Agenda

- Disclaimer..
- What format does my data need to be for the Machine Learning frameworks?
- Transforming Data
- Options/Alternatives
- Moving Forward..

© 2016 IBM Corporation

How do I get data into PowerAI?

© 2016 IBM Corporation

What format does my data need to be in for Machine Learning?

- Labeled data
  - Metadata (this is a dog, cat, etc.)
  - Sub folders named appropriately (/dog /cat)
  - CSV. (prevalent)
- TensorFlow
  - TFRecords
- Caffe
  - Blobs

© 2016 IBM Corporation

Transforming Data

ANIMAL	VALUE
	?
	?
	?
	?
	?

ANIMAL	DUCK?	BEAVER?	WALRUS?
	1	0	0
	0	1	0
	1	0	0
	0	0	1
	0	1	0

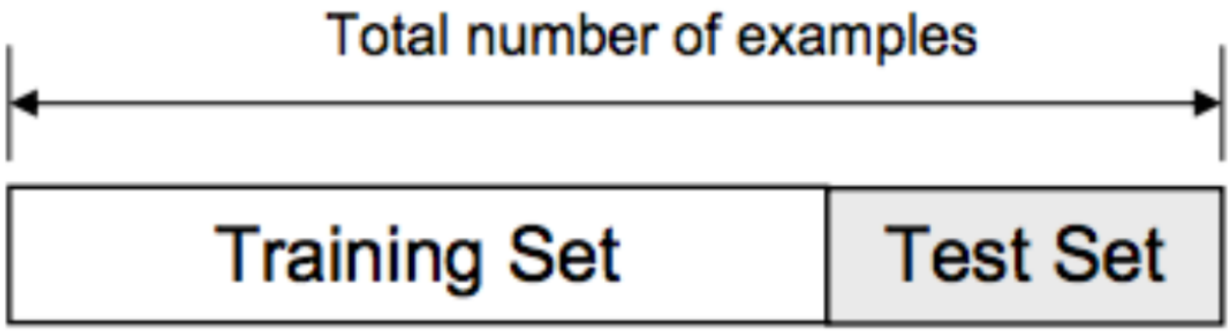
© 2016 IBM Corporation

Credit - Udacity

Transforming Data (null)

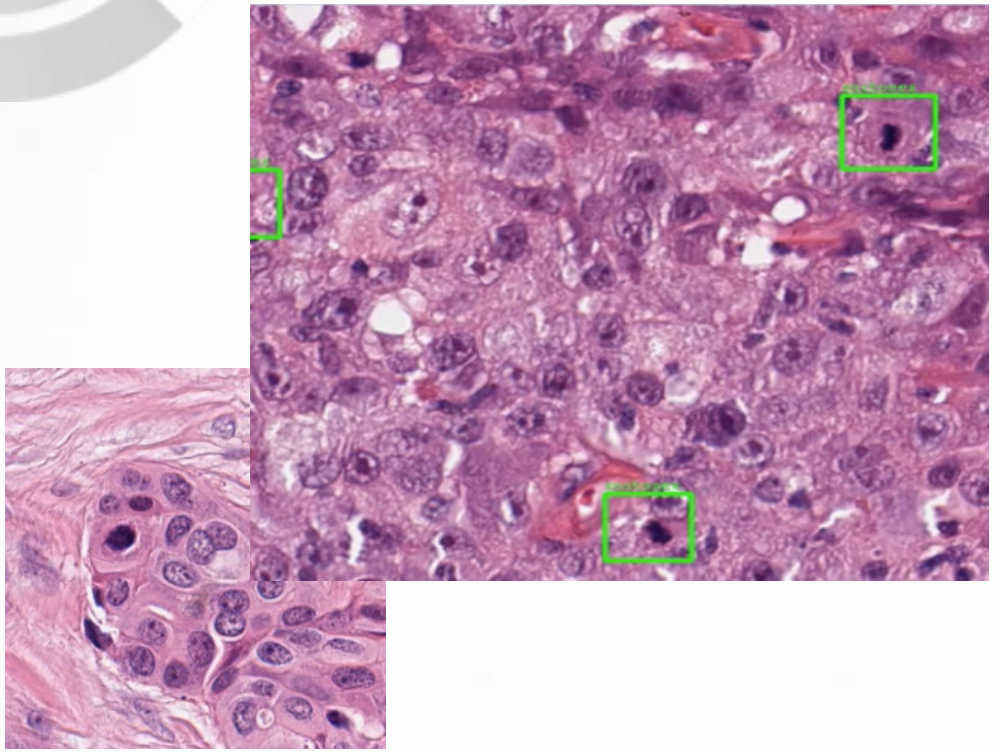
© 2016 IBM Corporation

Transforming Data



© 2016 IBM Corporation

Transforming Data



Tumor Proliferation Assessment – mitosis detection
Images from electron-microscope
Size of image - 70K * 60K

Data Transformation

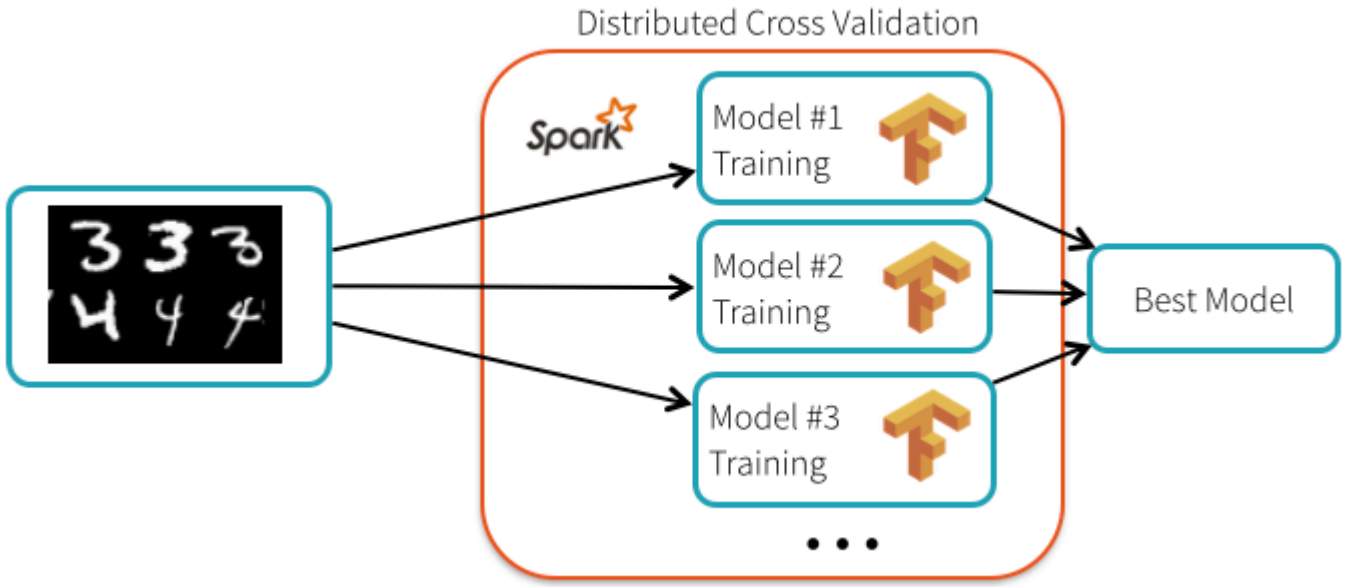
Framework	Format	Input Size (Faster R-CNN)
Caffe	LMDB	1K*1K
TensorFlow	TensorRecord	1K*1K

Data Distribution among training, validation and testing

Data Shuffle

© 2016 IBM Corporation

Options/Alternatives



© 2016 IBM Corporation

DSX

- Access data from local files
  - Local CSV data.
  - Loaded and transformed to DSX “Asset”
  - Drag and drop/file explorer
- Access HDFS data
  - Cloudera/Hortonworks/BigInsights
- Access RDB Data
  - Scala, Python, R APIs
  - Db2, Netezza, Informix, Oracle, Mongo
- Remote data?

```
import jaydebeapi

#Enter the values for you database connection
dsn_database = "BLUDB"            # e.g. "BLUDB" Name of the database
dsn_hostname = "dashdb.services.dal.bluemix.net" # e.g.: "bluemix05.bluforcloud.com"
dsn_port = "50000"                # e.g. "50000" Database port number
dsn_uid = "dashxxxx"        # e.g. "dash104434" User id
dsn_pwd = "xxxxxxx"       # e.g. "7dBZ3jWt9xN6$o0JiX!m" User password for the database

connection_string='jdbc:db2://'+dsn_hostname+':'+dsn_port+'/'+dsn_database+':user='+dsn_uid+
conn = jaydebeapi.connect("com.ibm.db2.jcc.DB2Driver", connection_string)

curs = conn.cursor()
curs.execute("select * from db2inst1.sales")
curs.fetchall()
```

© 2016 IBM Corporation

Thank You
Vinaka
köszönöm
Kiitos
Maake
Asante
Shukria
Dhanyavadagalu
감사합니다
Dankscheen
спасибо
Kam Sah Hammida
ارکش
Manana
Dankon
Diolch i Chi
Terima Kasih
Matondo
Dank Je
Mauruuru
Biyan
Taiku
Blagodaram
Ngiyabonga
Dziekuje
Chokrane
Tack
Grazie
Juspaxar
Arigato
Mochchakkeram
Gracias
நன்றி
Bedankt
D'akujem
धन्यवाद
cảm ơn bạn
Khap
Paldies
Tingki
Ua Tsaug Rau Koj
Grazas
Gratias Tibi
Děkuji
Nirringrazzjak
Di Ou Mèsi
Kia Ora
Kop Khun
Obrigado
Suksama
Rahmat
Nuwun
Hvala
ありがとう
谢谢
хвала
Welalin
Djiere Dieuf
Misaotra
Matur
Danke
Merci
Go Raibh Maith Agat
Eskerrik Asko
Salamat
ขอขอบคุณคุณ
Najis Tuke