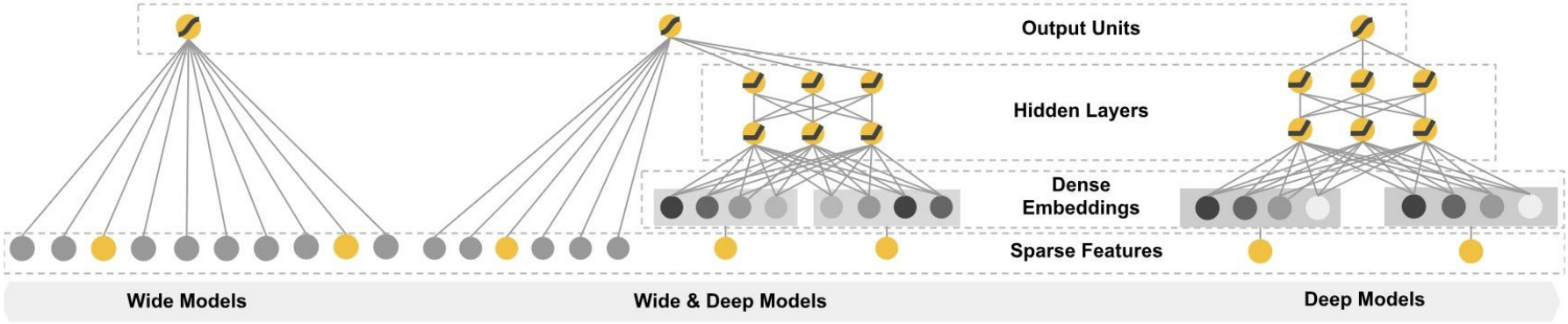
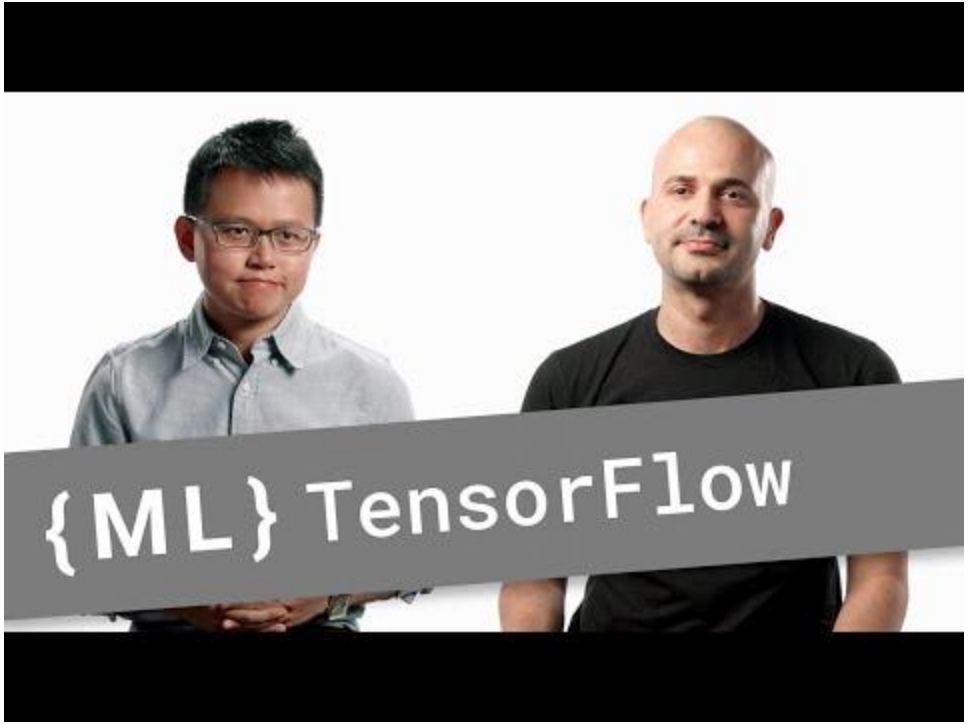
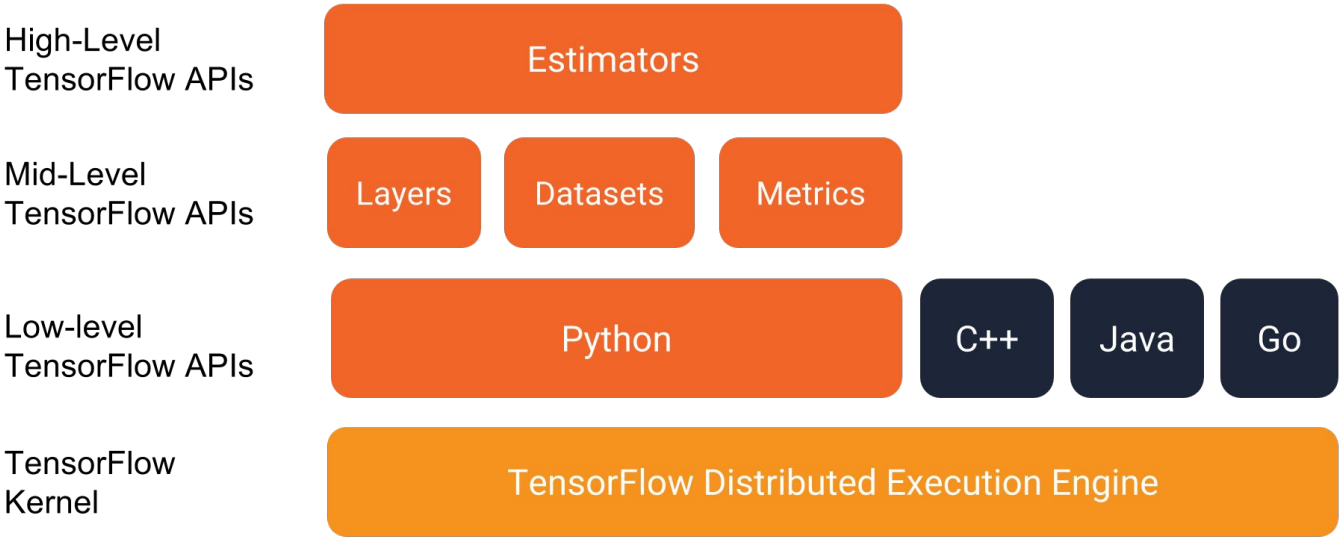
Practical  - Wide & Deep Learning Workshop

Thank you for the support!

What is Wide and Deep Learning?




{ML} TensorFlow


High-Level
TensorFlow APIs
Estimators
Mid-Level
TensorFlow APIs
Layers
Datasets
Metrics
Low-level
TensorFlow APIs
Python
C++
Java
Go
TensorFlow
Kernel
TensorFlow Distributed Execution Engine

Layer Architecture

The Wide Model

```
network = tflearn.fully_connected(network, n_inputs, activation="linear",
name="wide_linear", bias=False) # x*W (no bias)


network = tf.reduce_sum(network, 1, name="reduce_sum")      # batched
sum, to produce logits
network = tf.reshape(network, [-1, 1])
```

The Deep Model

```
n_nodes=[100, 50]
network = tf.concat(1, [wide_inputs] + flat_vars, name="deep_concat")
for k in range(len(n_nodes)):
    network = tflearn.fully_connected(network, n_nodes[k], activation="relu",
name="deep_fc%d" % (k+1))
network = tflearn.fully_connected(network, 1, activation="linear",
name="deep_fc_output", bias=False)
```

The Wide & Deep model

$$P(Y = 1|\mathbf{x}) = \sigma(\mathbf{w}_{wide}^T[\mathbf{x}, \phi(\mathbf{x})] + \mathbf{w}_{deep}^T a^{(l_f)} + b)$$

Census Income

Data Set Characteristics: Multivariate

Attribute Characterictics: Categorical, Integer

Associated Tasks: Classification

Number of Instances: 48842

Number of Attributes: 14

Missing Values: Yes

Area: Social

Date Donated: 1996/05/01

Number of Web Hits: 265651

URL: https://archive.ics.uci.edu/ml/datasets/Census+Income

Census Income

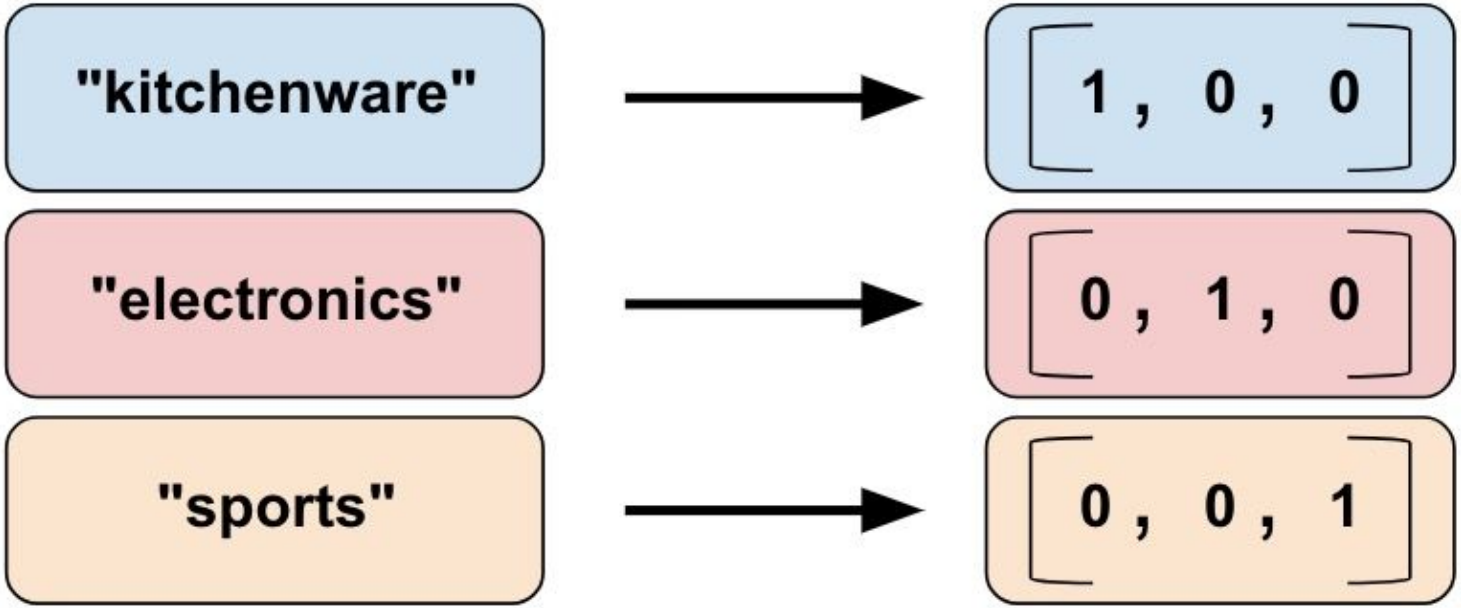
Attribute Information:
Listing of attributes:

>50K, <=50K.

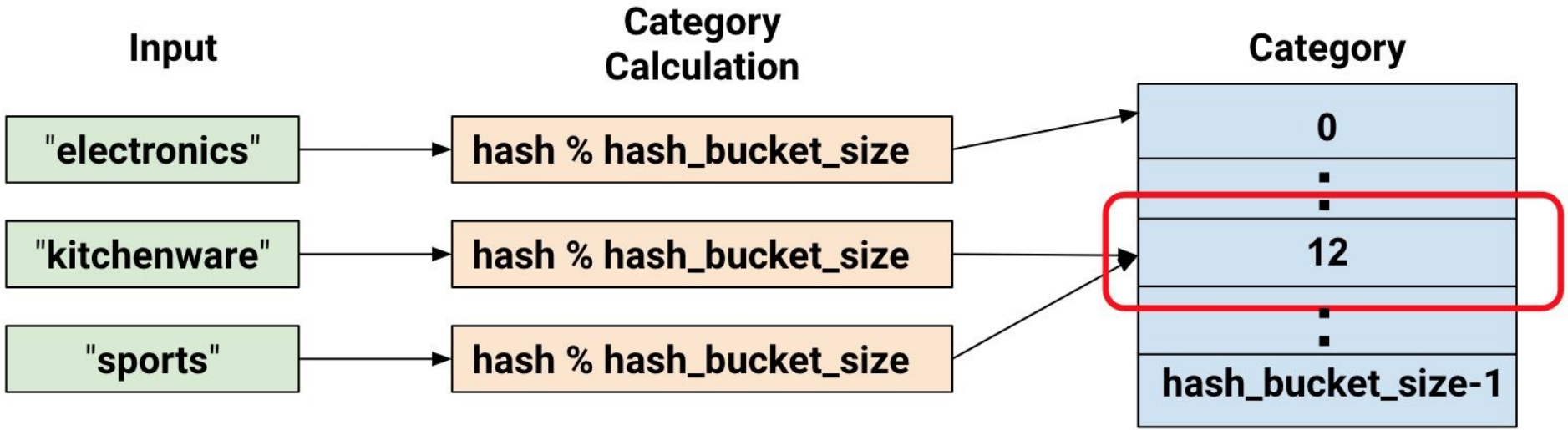
age: continuous.
workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
fnlwgt: continuous.
education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
education-num: continuous.
marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
sex: Female, Male.
capital-gain: continuous.
capital-loss: continuous.
hours-per-week: continuous.
native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

Let’s write some code

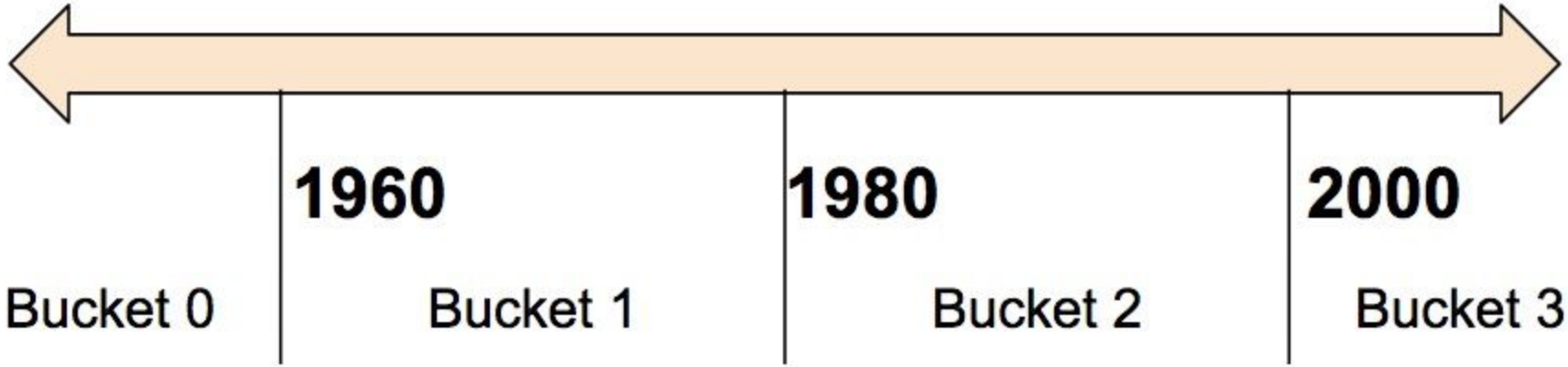
Categorical Columns with Vocabulary List



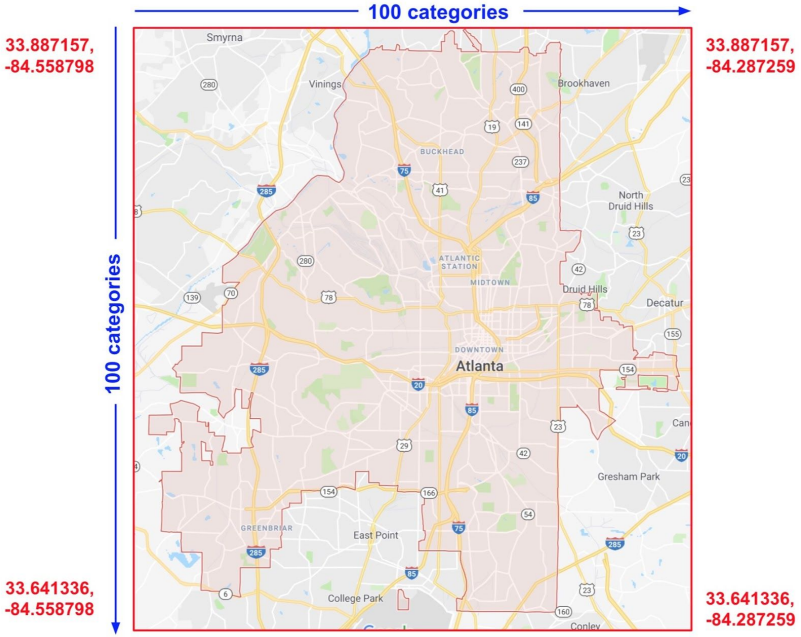
Categorical Columns with Hash Bucket



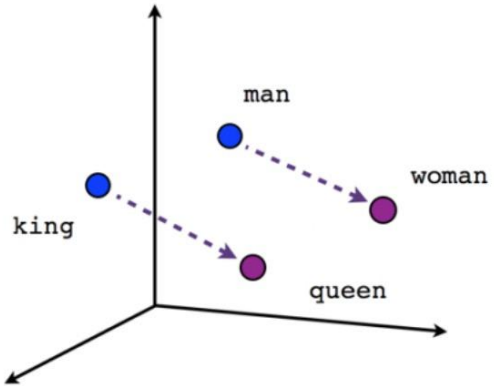
Bucketized Columns



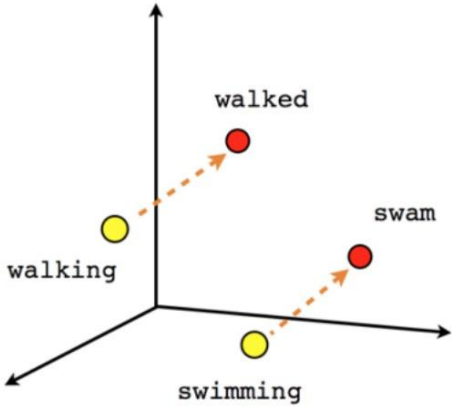
Crossed Feature Columns



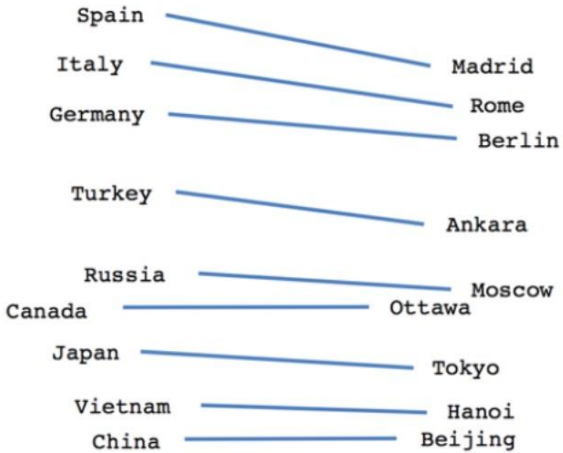
Embeddings



Male-Female



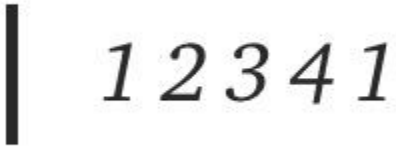
Verb tense



Country-Capital

Embeddings

Embeddings



Embeddings

Indices	Latent Factors					
1	.32	.02	.48	.21	.56	.15
2	.65	.23	.41	.57	.03	.92
3	.45	.87	.89	.45	.12	.01
4	.65	.21	.25	.45	.78	.82

Embedding Matrix

Thank you

Daniel Kuchta
Head of Machine Learning, GlobalLogic Slovakia

daniel.kuchta@globallogic.com

@danonops

@gama843