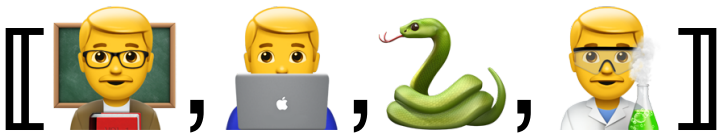
Lecture Notes for

Machine Learning in Python



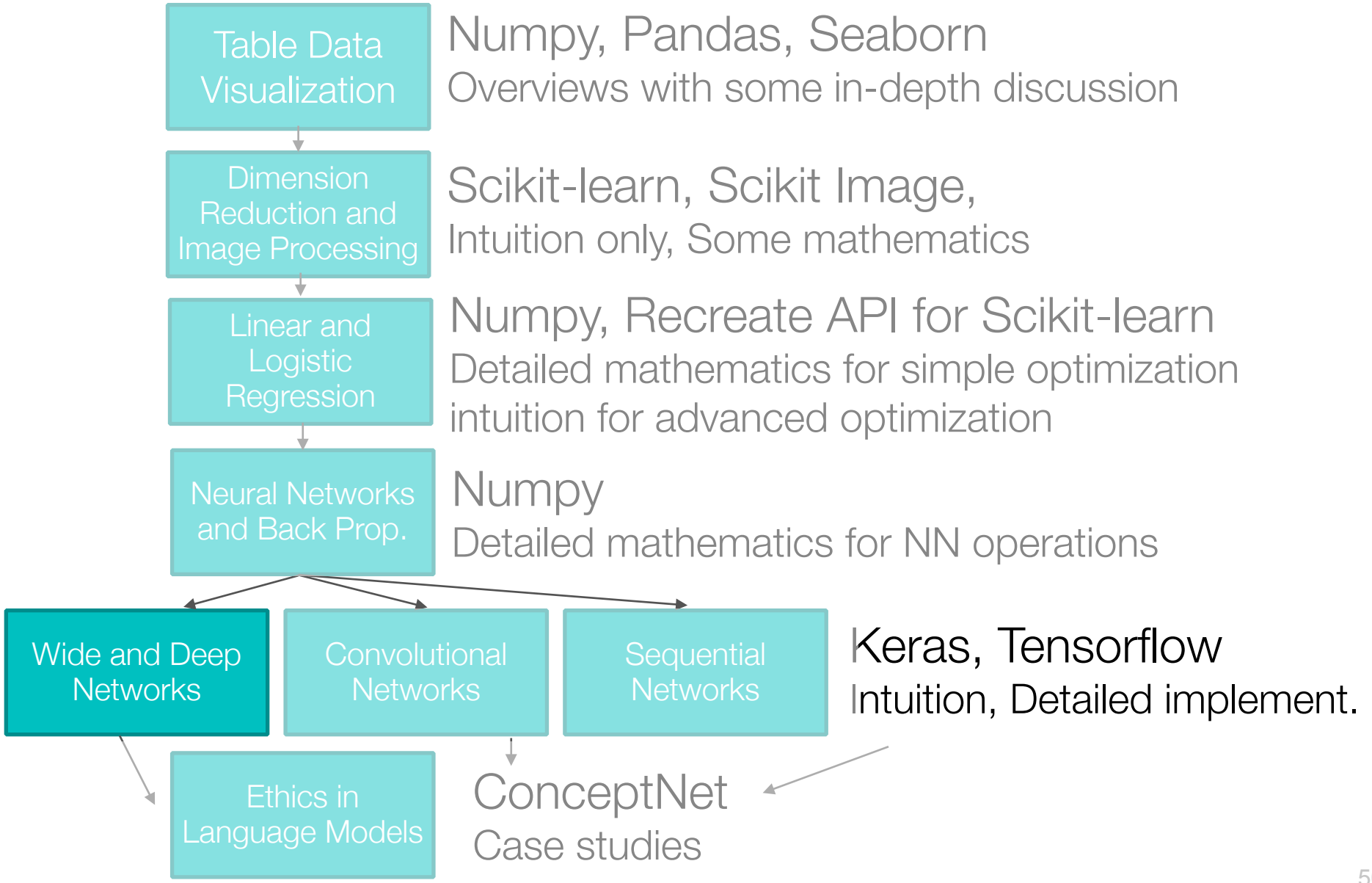
Professor Eric Larson

Wide and Deep Networks

Logistics and Agenda

- Logistics
  - Wide/Deep due soon!
  - Lecture video
  - A brief intro to career opportunities (Sonjia Stafford)
- Agenda
  - Wide/Deep Finish Demo
  - Town Hall
  - (if time) Basic CNN architectures

Class Overview, by topic



Review: Feature Embeddings

Trainable Matrix, $\mathbf{W}^{(emb)}$, in Network

Reduced Dimensions

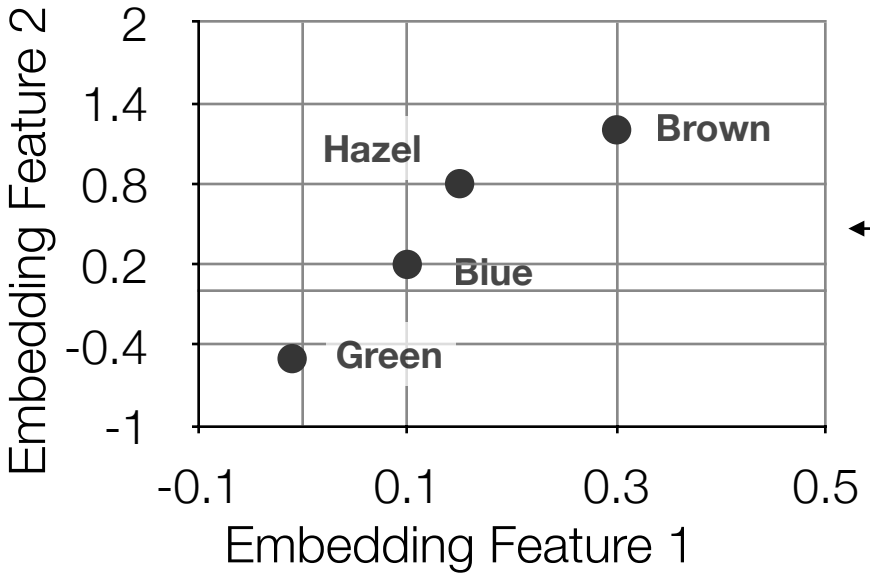
0.1	0.3	-0.01	0.15
0.2	1.2	-0.5	0.8

Num Categories

Eye Color

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{matrix} \text{Blue} \rightarrow 0 \\ \text{Brown} \rightarrow 1 \\ \text{Green} \rightarrow 2 \\ \text{Hazel} \rightarrow 3 \end{matrix}$$

High Dimensional Categorical Data Space

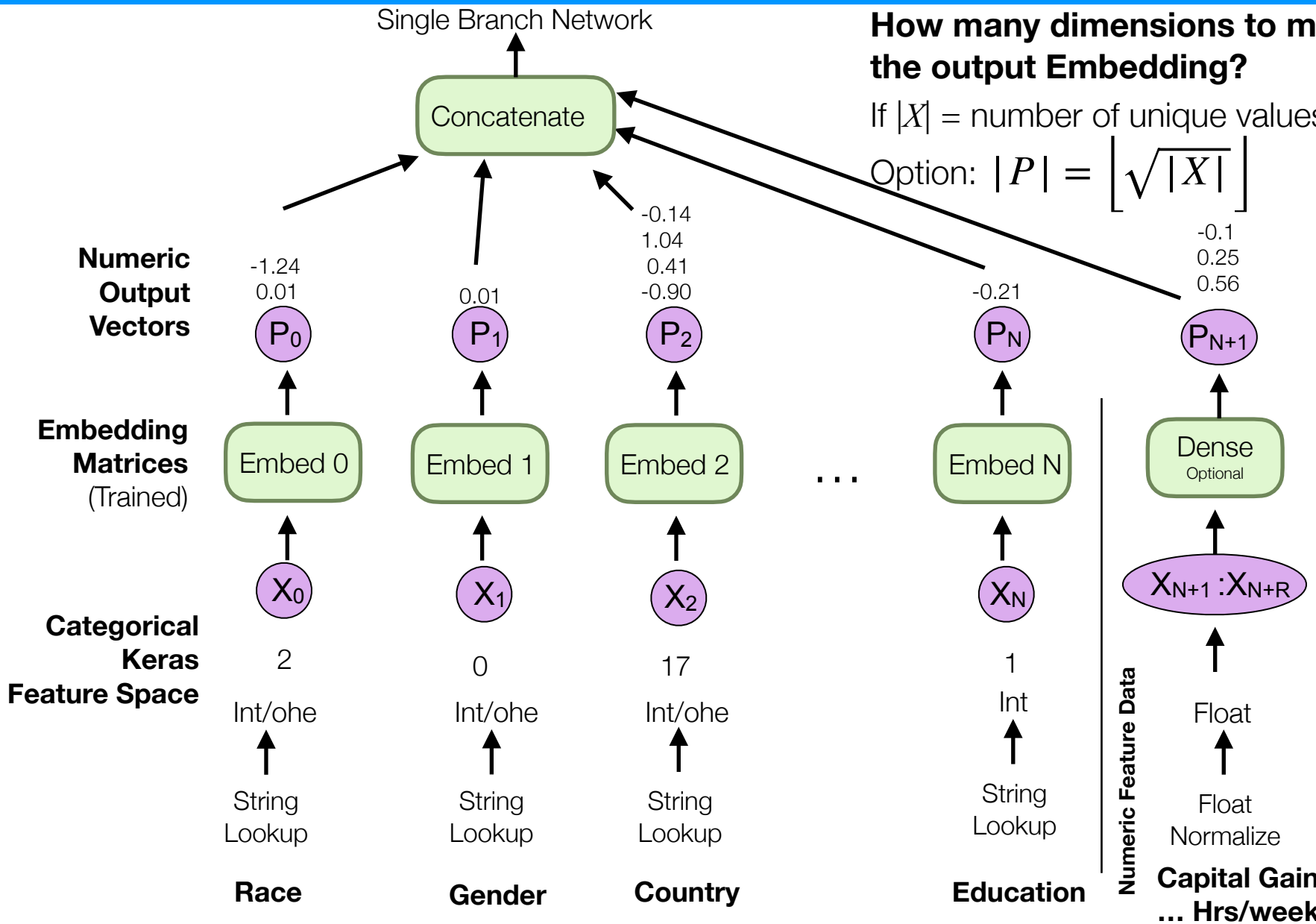


Lower Dimensional Embedding Space

Embedding matrix

$$\mathbf{a} = \mathbf{W}^{(emb)} \cdot \mathbf{x}_{OHE}$$

Review: Embeddings with Feature Spaces



Self Test:

- Keras `FeatureSpaces` …
  - A. … provide operations in the tensorflow computation graph
  - B. … can be used to create intermediate `tf.Dataset` representations.
  - C. … require the call to `.adapt` to learn from an example dataset about aspects like vocabulary size and normalization values
  - D. All of the Above

Finish Demo

Adding Embedding Branches

`10a. Keras Wide and Deep as TFData.ipynb`

Wide and Deep Networks

Wide and Deep

Wide & Deep Learning for Recommender Systems

Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah

Google Inc.*

ABSTRACT

Generalized linear models with nonlinear feature transfor-

have never or rarely occurred in the past. Recommenda-tions based on memorization are usually more topical and

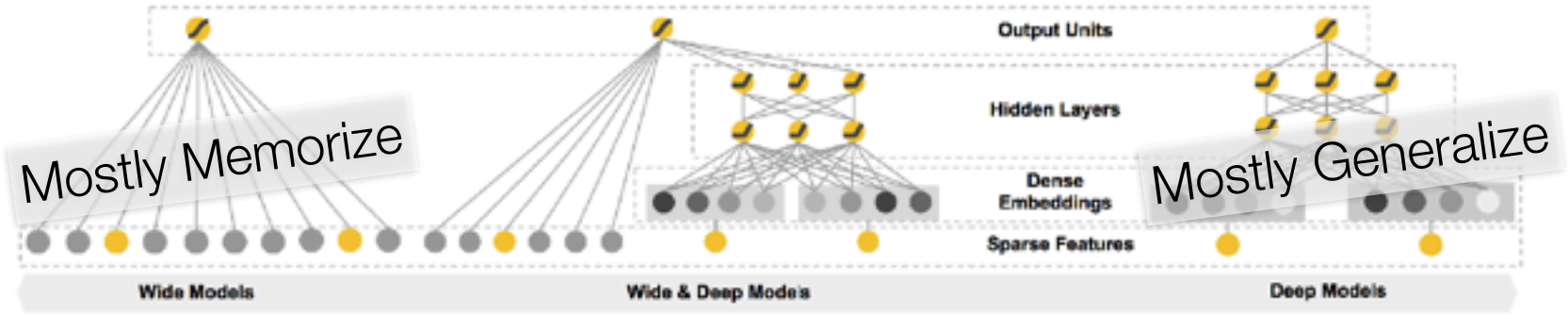
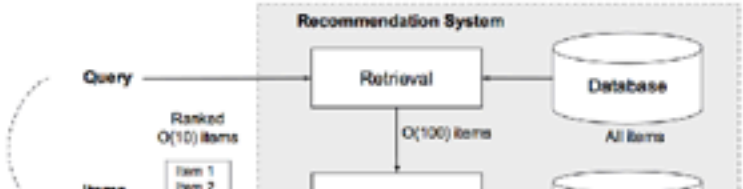


Figure 1: The spectrum of Wide & Deep models.

linear model with feature transformations for generic recommender systems with sparse inputs.

- The implementation and evaluation of the Wide & Deep recommender system productionized on Google



Why wide and deep?

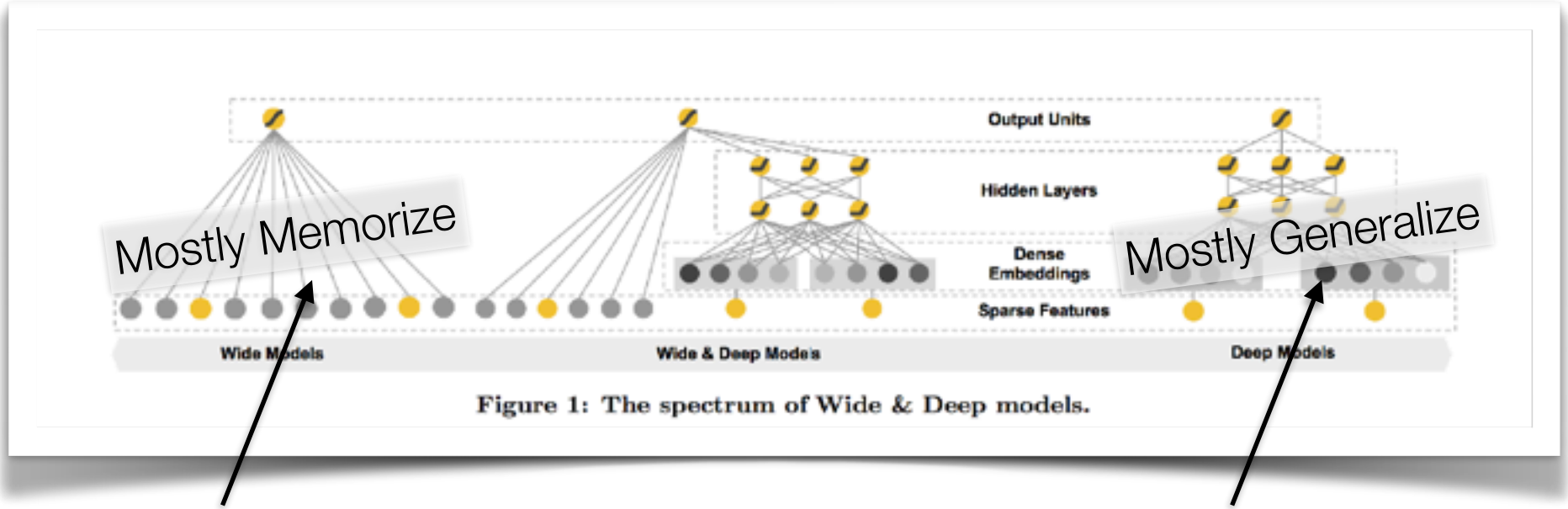


Figure 1: The spectrum of Wide & Deep models.

But why memorize?

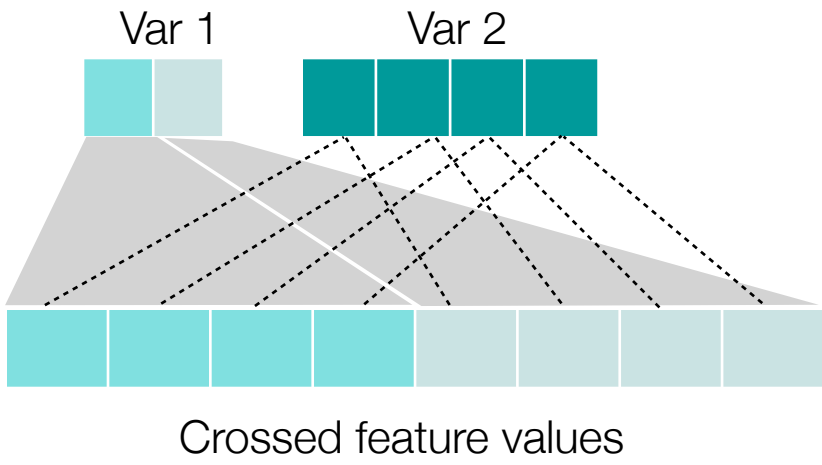
Obvious!

- Categorical values have combinations that repeat!
  - so memorizing these values is not necessarily a bad strategy
  - let's make memorizing easy on one network

Wide networks (Memorize?)

- Wide refers to the expansion of features set
- Crossed feature columns of categorical features

- Movie Rating
  - G
  - PG
  - PG-13
  - R
  - Else
- Movie Genre
  - Action
  - Drama
  - Comedy
  - Horror
  - Else

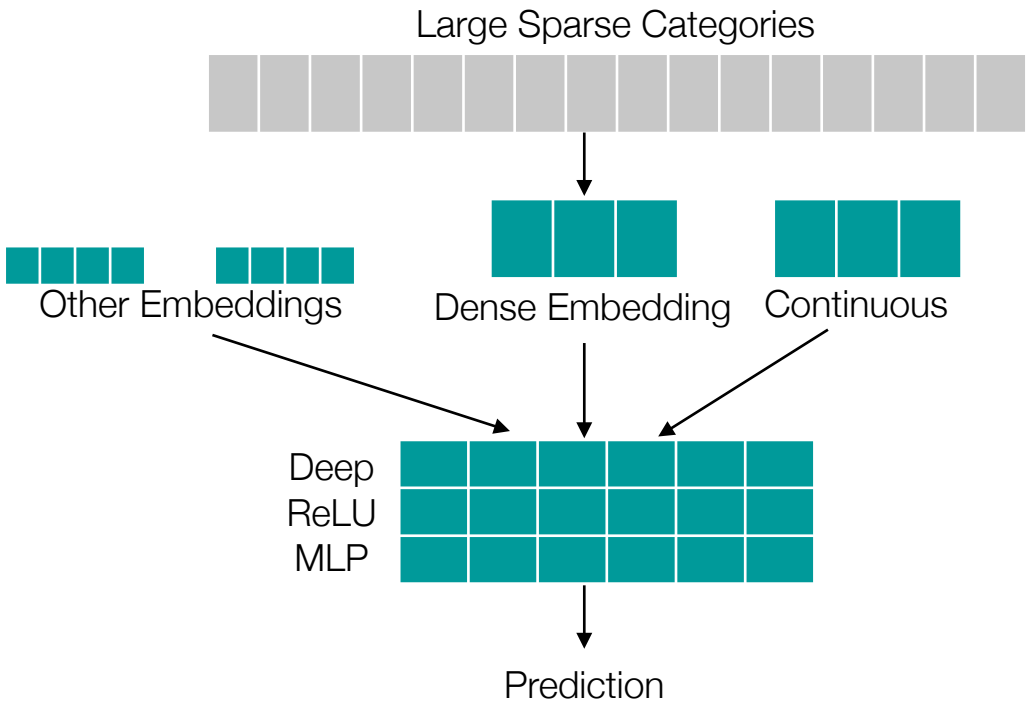


- Crossed feature "Rating-Genre"
  - G-Action, G-Drama, G-Comedy, G-Horror, G-else
  - PG-Action, PG-Drama, PG-Comedy, PG-Horror, G-else
  - and so on … one hot encoded

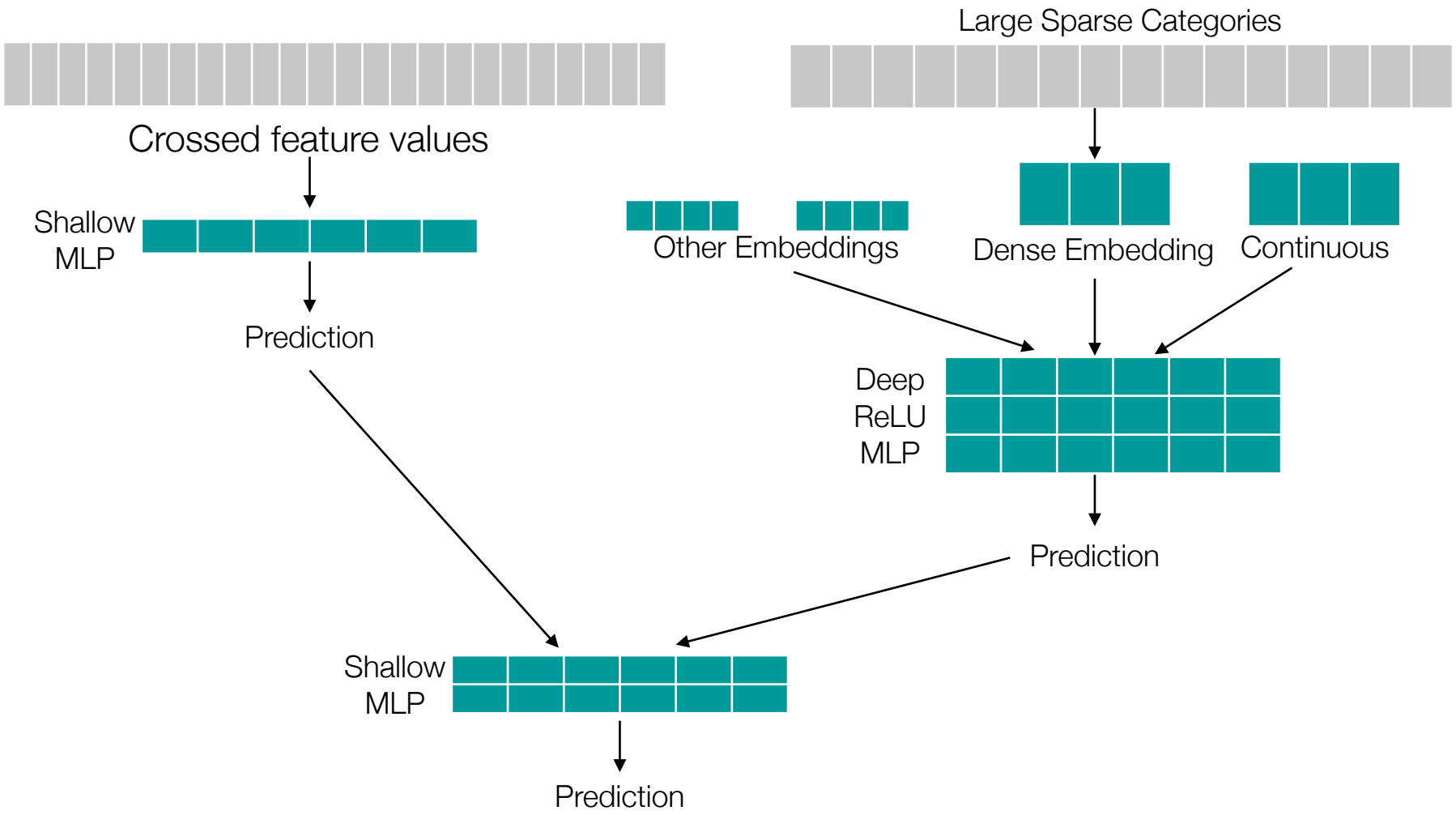
Deep Features: What we have already done!

- Deep refers to increasingly narrow hidden layers
- **Essentially the same as what we already did!**

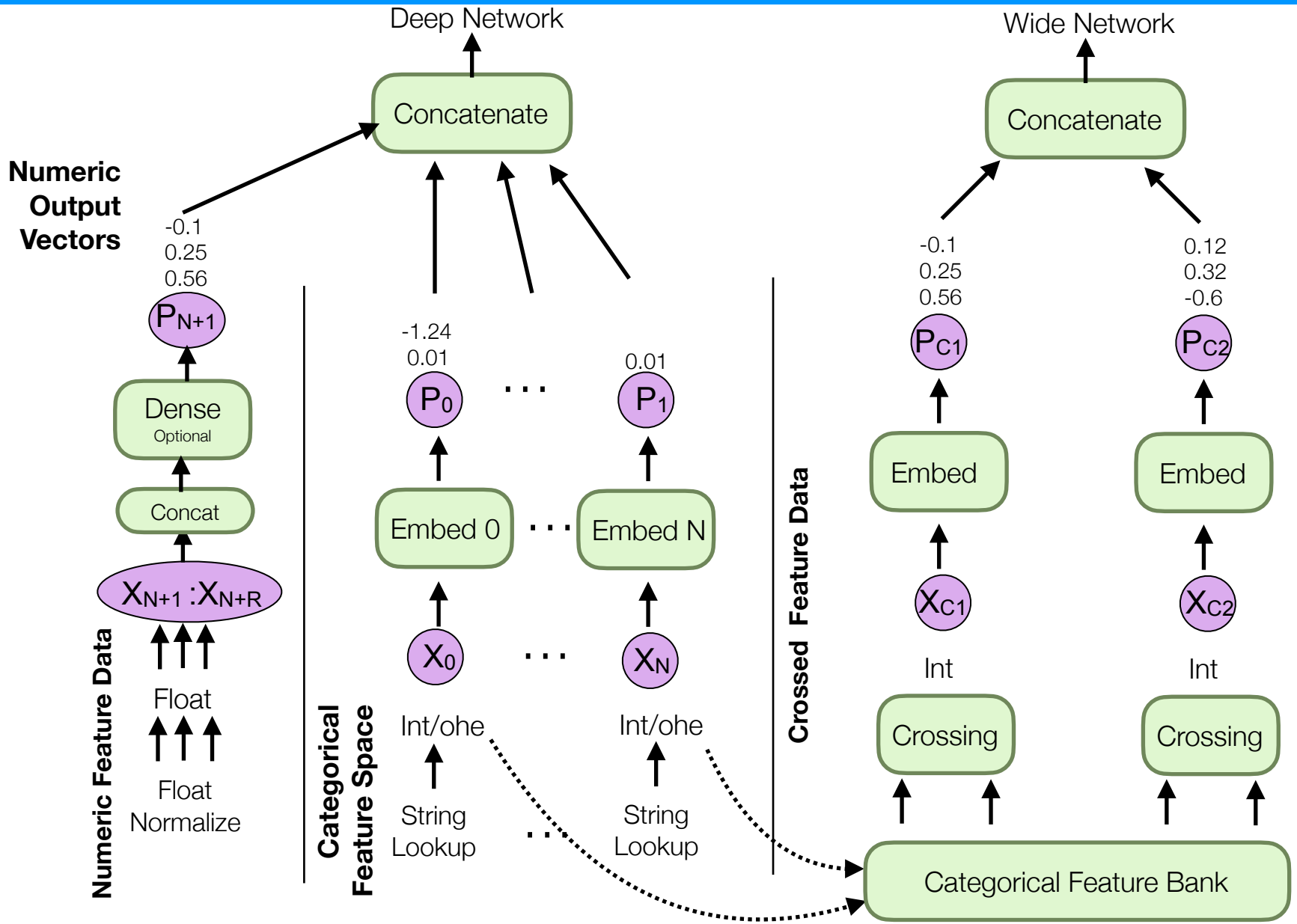
- Movie Actors
  - Armand Assante
  - Meryl Streep
  - Danny Trejo
  - Kevin Bacon
  - Audrey Hepburn
  - …
-



Combining Memorization and Generalization



Crossed Embeddings in Keras



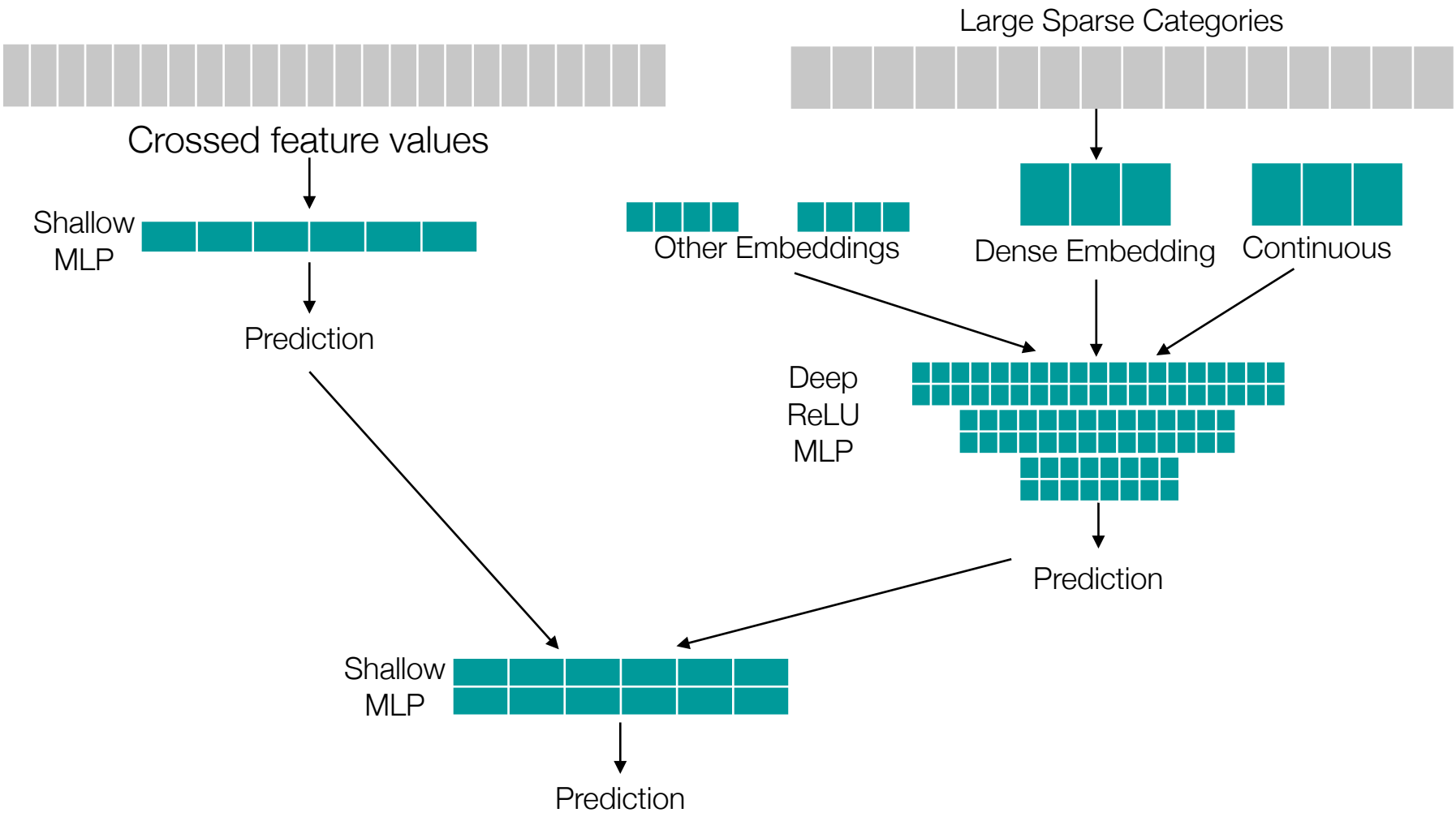
Wide and Deep

Demo

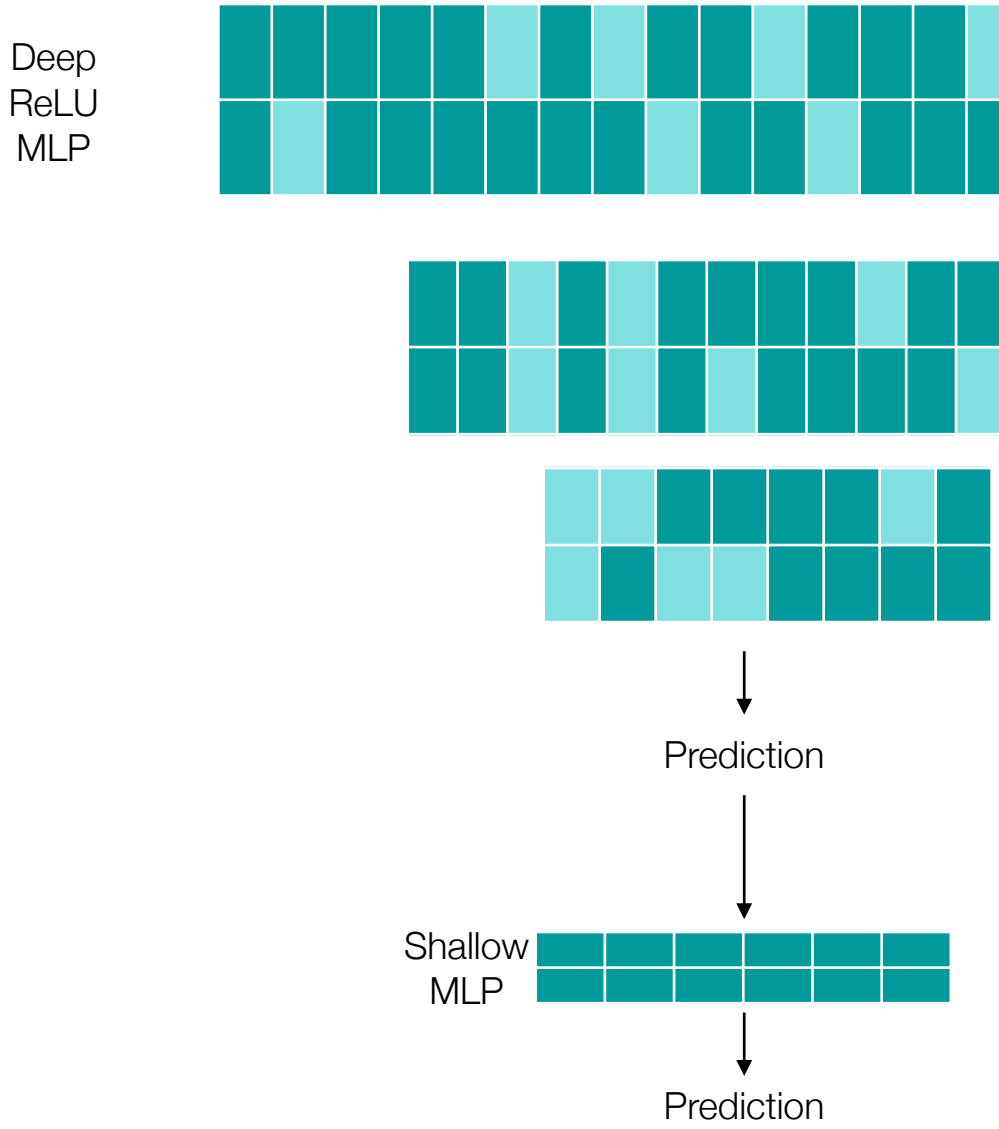
Adding Wide Branches

`10a. Keras Wide and Deep as TFData.ipynb`

Dropout: Helping prevent over fitting…



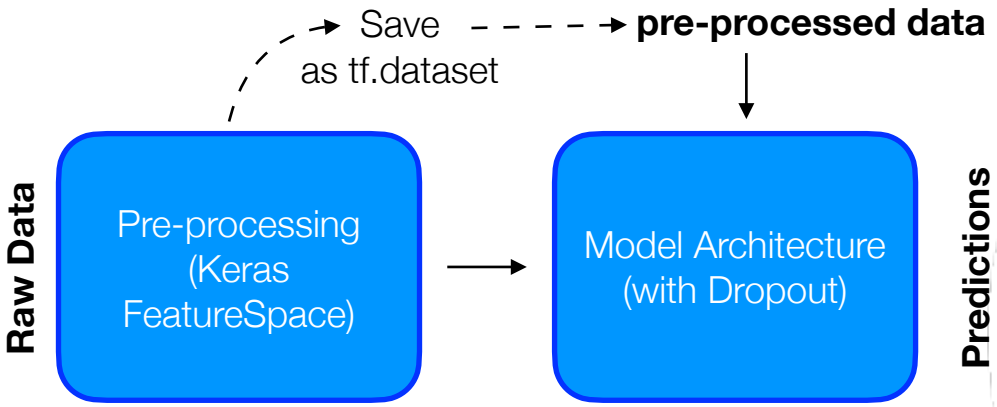
Dropout: Helping prevent over fitting…

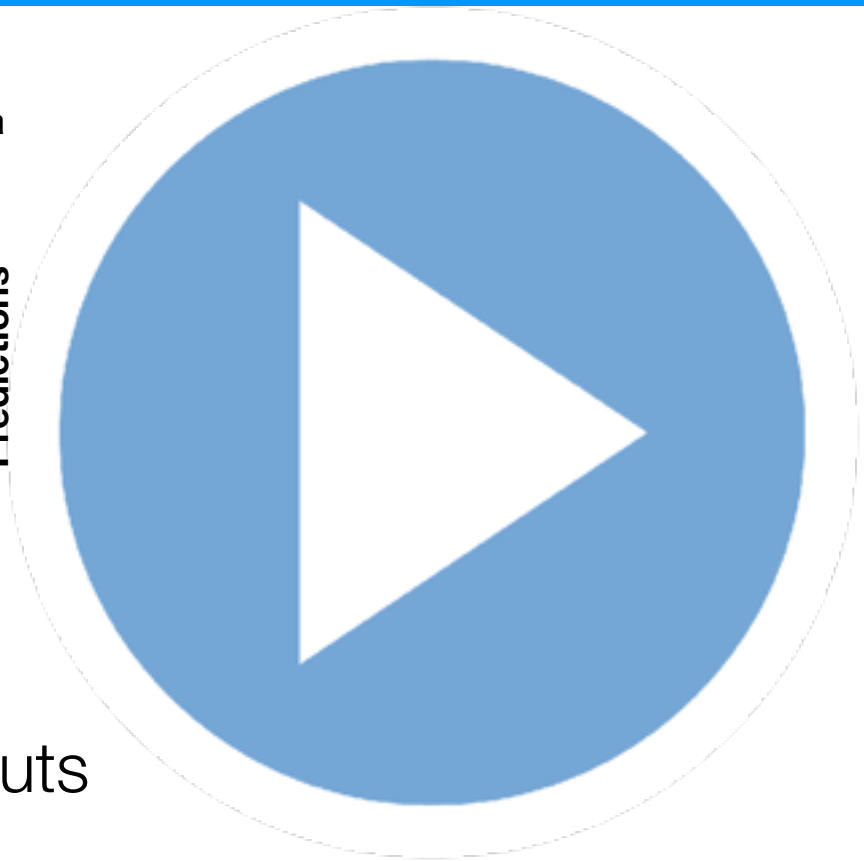


- Do not update all weights for each batch.
- Randomly choose which weights to update
- Choose different set of weights for each batch (random percentage)
- **Caveat**: does extend required training time for convergence

Wide and Deep

Demo

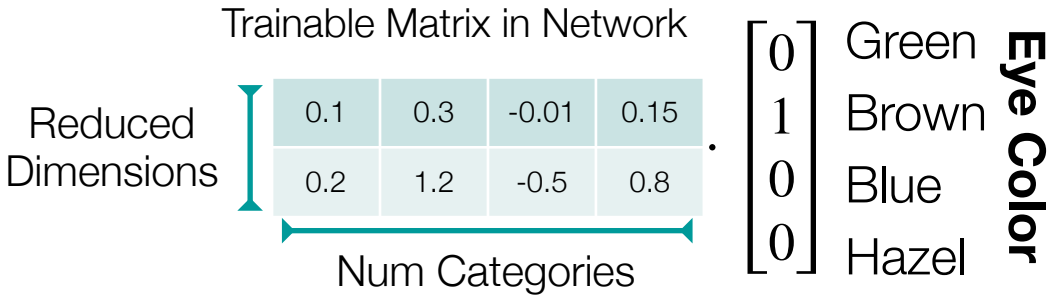


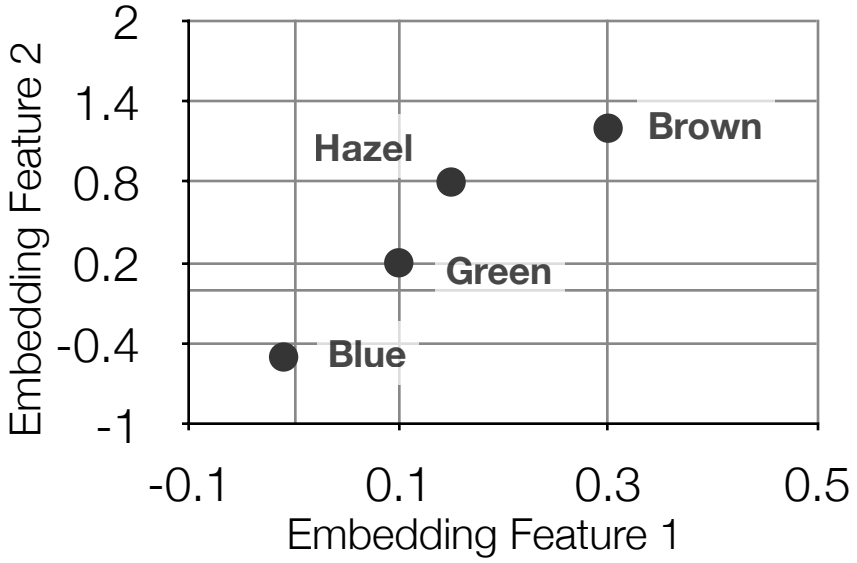


Additional tuning with Dropout
Using saved pre-processed outputs

`10a. Keras Wide and Deep as TFData.ipynb`

Town Hall, Wide and Deep Networks





End of Session

- Next Time (or now if time):
  - Convolutional Neural Networks