Agenda

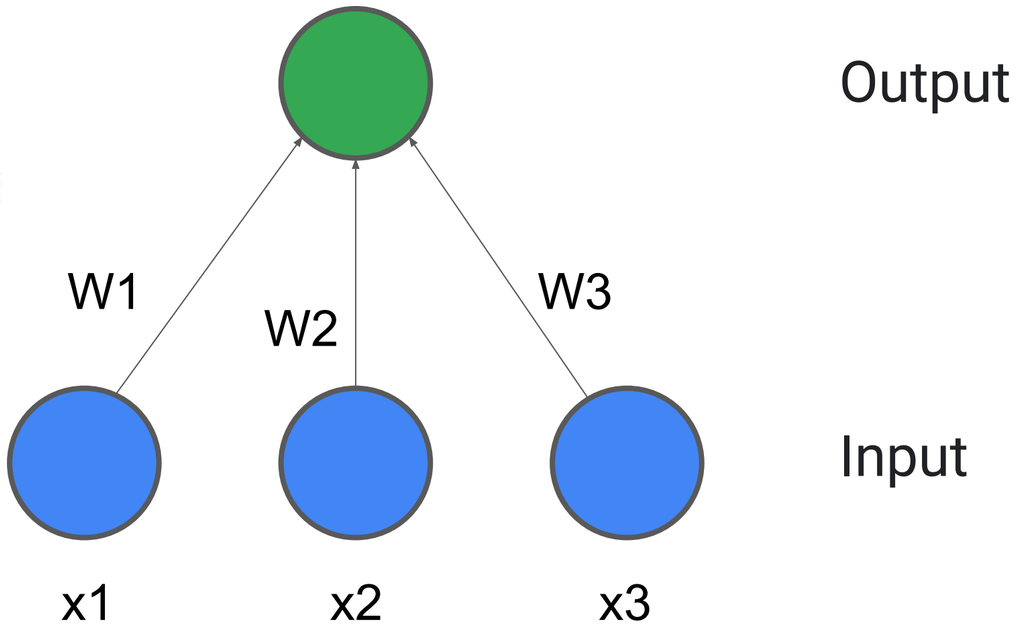
- Activation Functions
- Neural Networks with TF 2 and Keras
- Regularization

Agenda

- **Activation Functions**
- Neural Networks with TF 2 and Keras
- Regularization

A Linear Model can be represented as nodes and edges

$$output = W_1 * x_1 + W_2 * x_2 + W_3 * x_3$$



Add Complexity: Non-Linear?

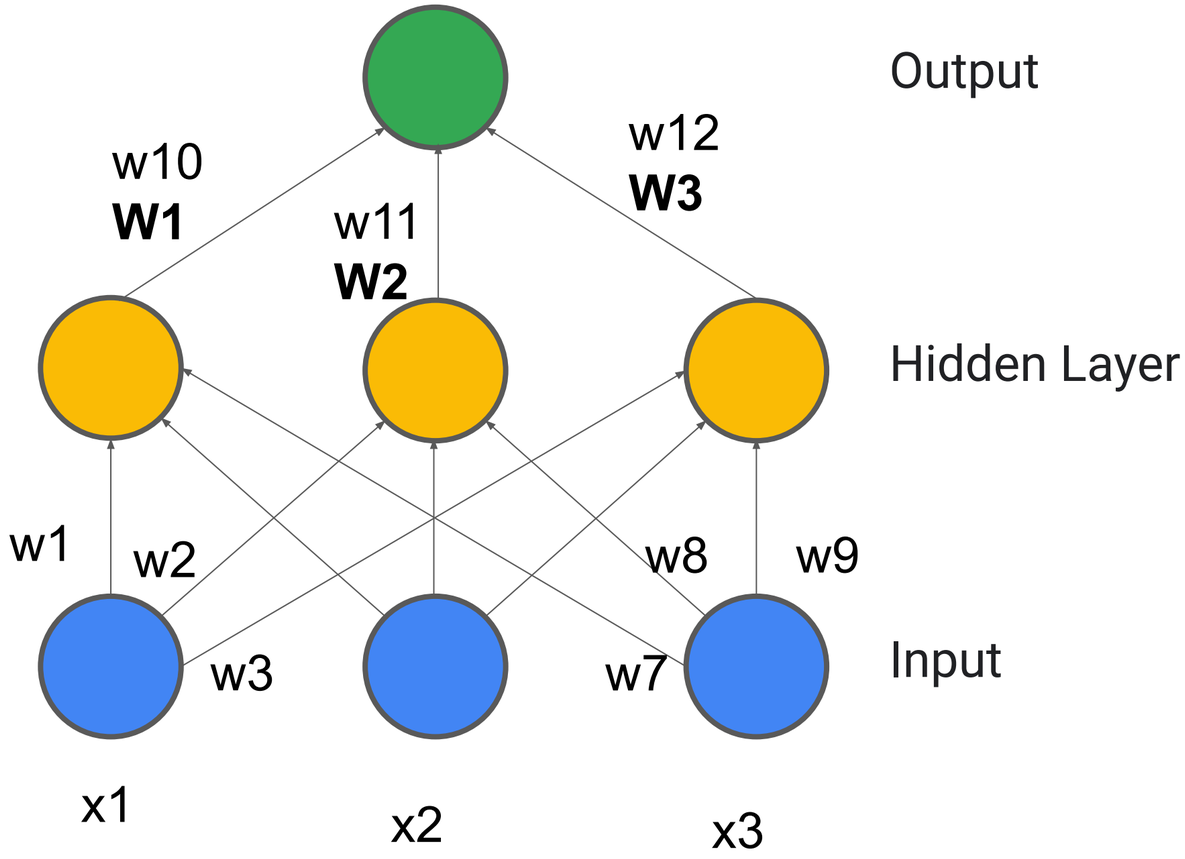
$$output = w_{10} * h_1 + w_{11} * h_2 + w_{12} * h_3$$

((w10 * w1) + (w10 * w4) + (w10 * w7)) * x1

+ ((w11 * w1) + (w11 * w4) + (w11 * w7)) * x2

+ ((w12 * w1) + (w12 * w4) + (w12 * w7)) * x3

$$= W_1 * x_1 + W_2 * x_2 + W_3 * x_3$$



Add Complexity: Non-Linear?

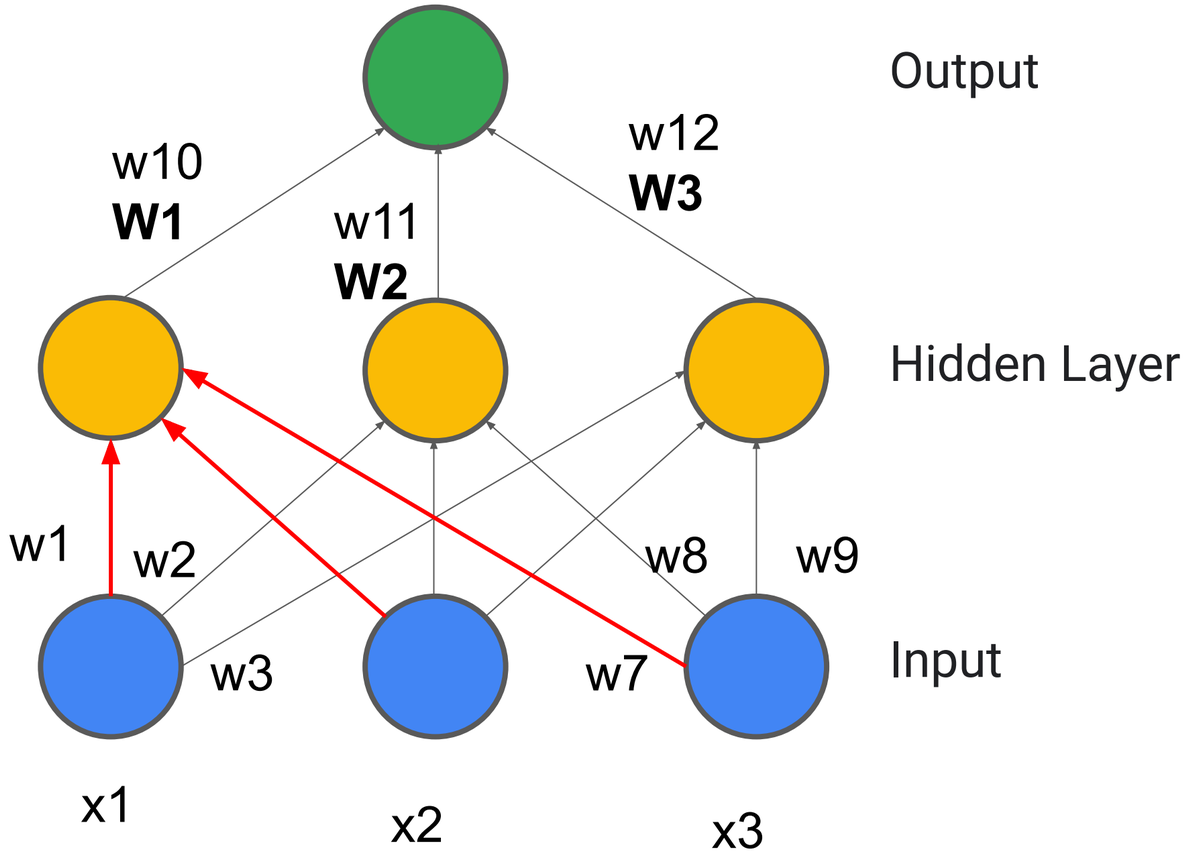
$$output = w_{10} * h_1 + w_{11} * h_2 + w_{12} * h_3$$

((w10 * w1) + (w10 * w4) + (w10 * w7)) * x1

+ ((w11 * w1) + (w11 * w4) + (w11 * w7)) * x2

+ ((w12 * w1) + (w12 * w4) + (w12 * w7)) * x3

$$= W_1 * x_1 + W_2 * x_2 + W_3 * x_3$$



Add Complexity: Non-Linear?

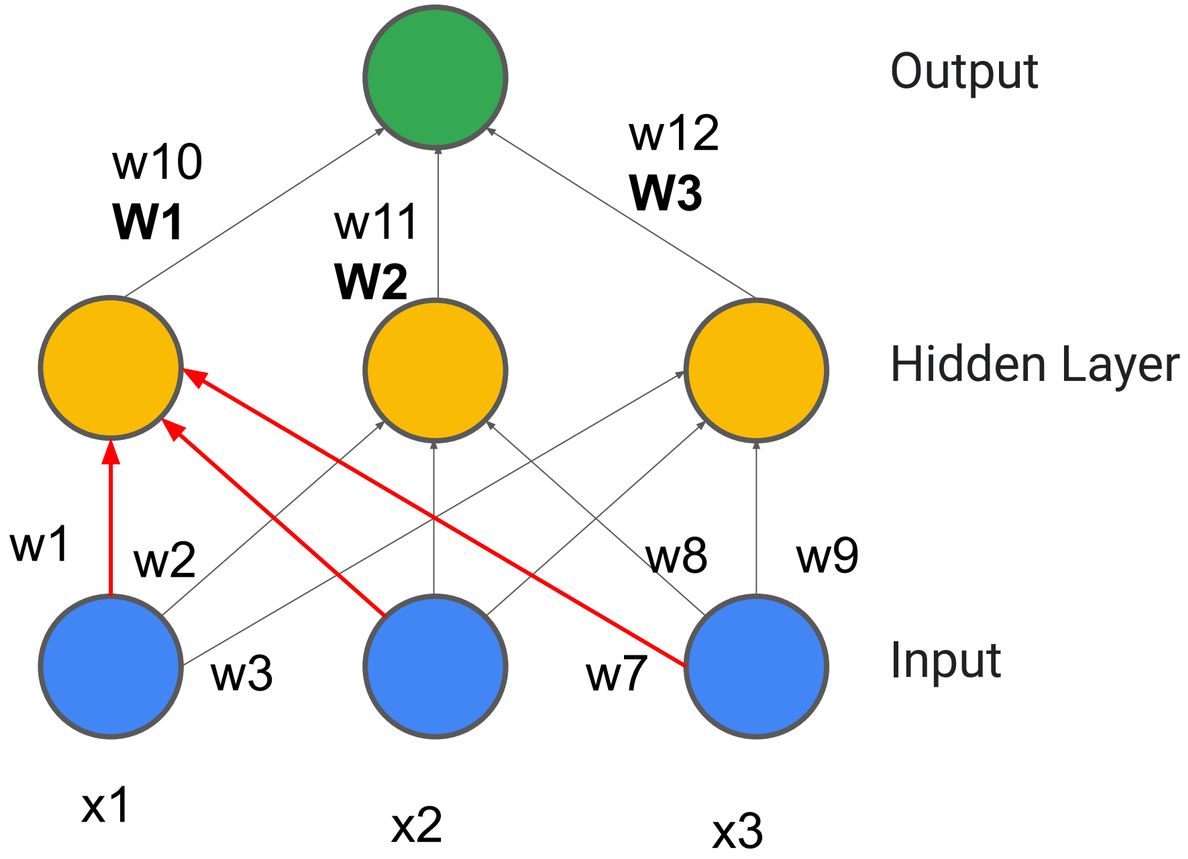
$$output = w_{10} * h_1 + w_{11} * h_2 + w_{12} * h_3$$

((w10 * **w1**) + (w10 * **w4**) + (w10 * **w7**)) * x1

+ ((w11 * w1) + (w11 * w4) + (w11 * w7)) * x2

+ ((w12 * w1) + (w12 * w4) + (w12 * w7)) * x3

$$= W_1 * x_1 + W_2 * x_2 + W_3 * x_3$$



Add Complexity: Non-Linear?

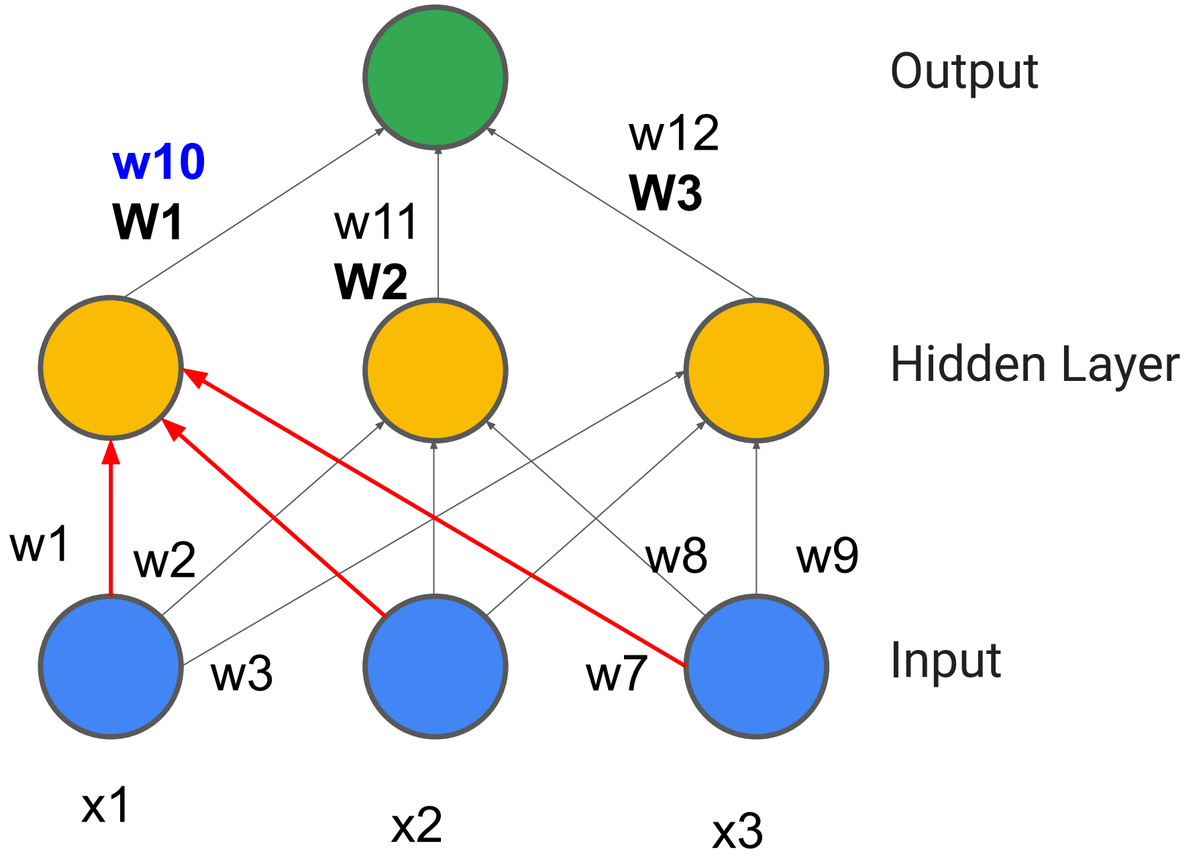
$$output = w_{10} * h_1 + w_{11} * h_2 + w_{12} * h_3$$

((**w10** * **w1**) + (**w10** * **w4**) + (**w10** * **w7**)) * x1

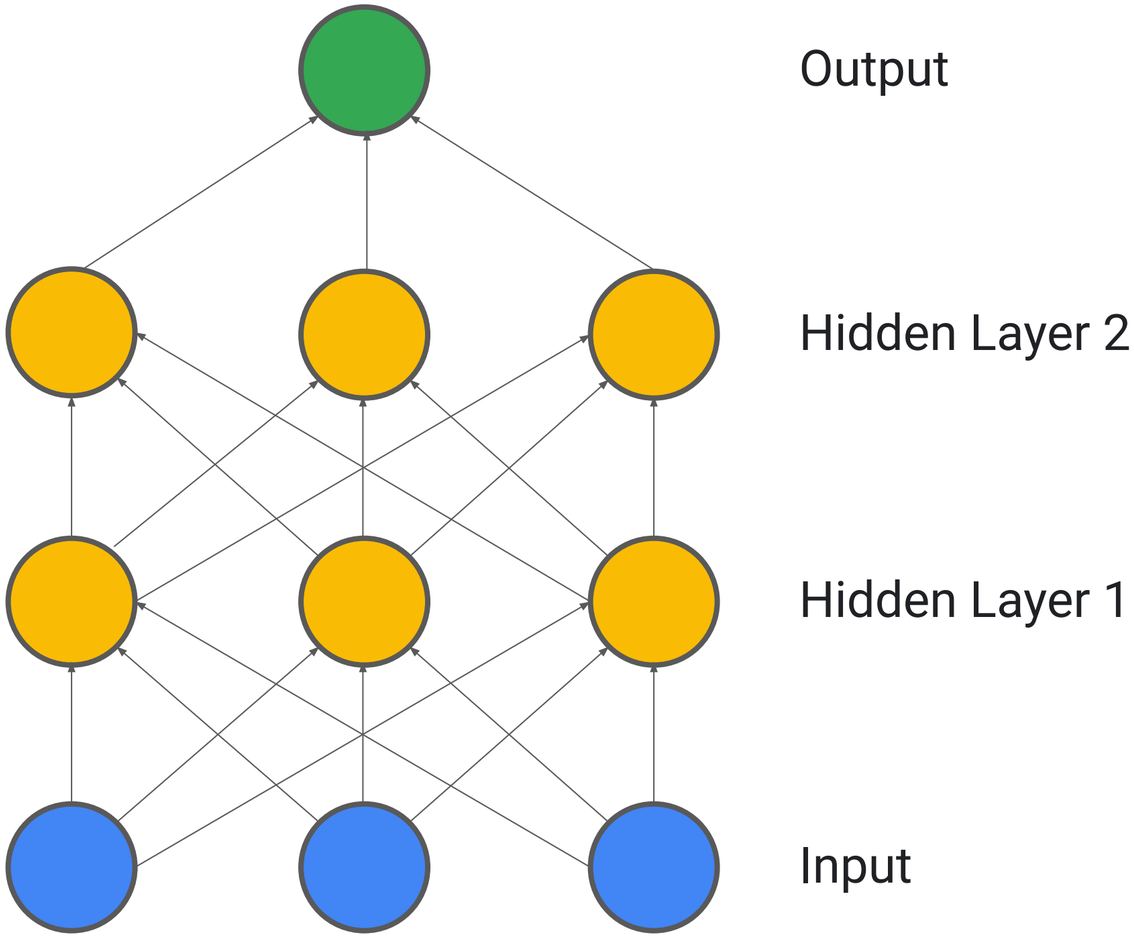
+ ((w11 * w1) + (w11 * w4) + (w11 * w7)) * x2

+ ((w12 * w1) + (w12 * w4) + (w12 * w7)) * x3

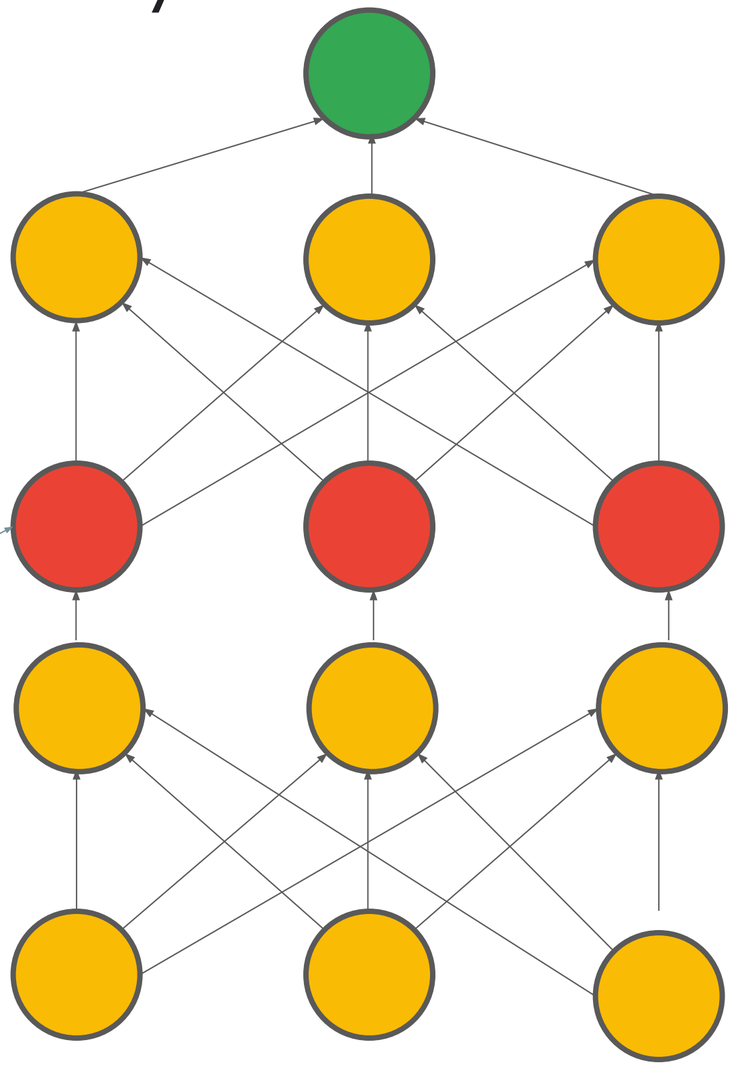
$$= W_1 * x_1 + W_2 * x_2 + W_3 * x_3$$



Add Complexity: Non-Linear?



Adding a Non-Linearity



Output

Hidden Layer 2

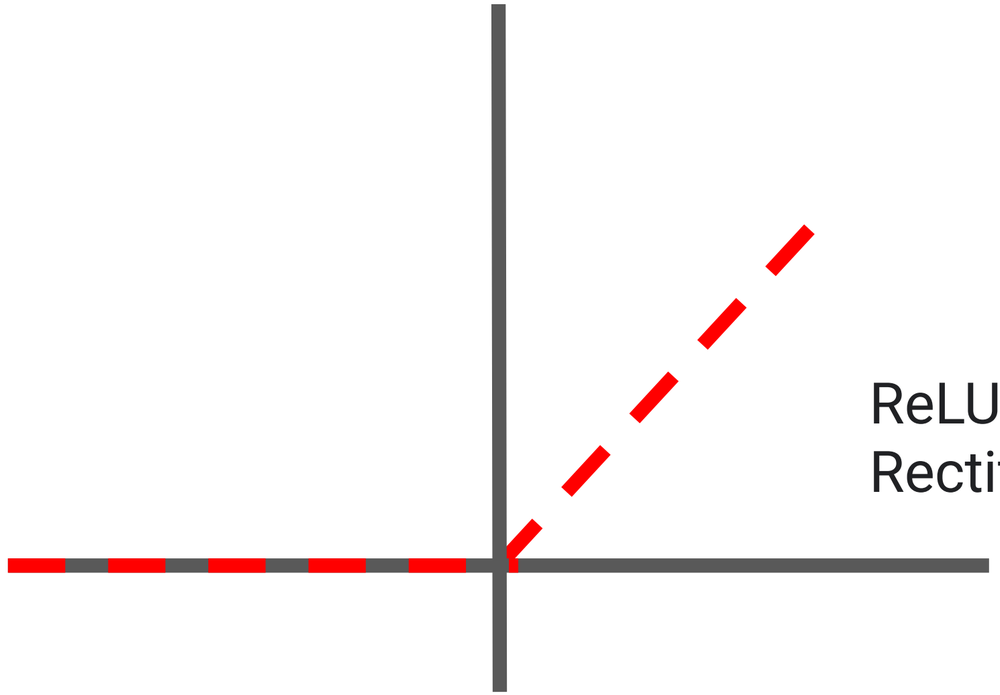
Non-Linear
Transformation Layer
aka Activation Function

Hidden Layer 1

Input

We usually don't draw Non-Linear Transforms.

Our favorite non-linearity is the Rectified Linear Unit

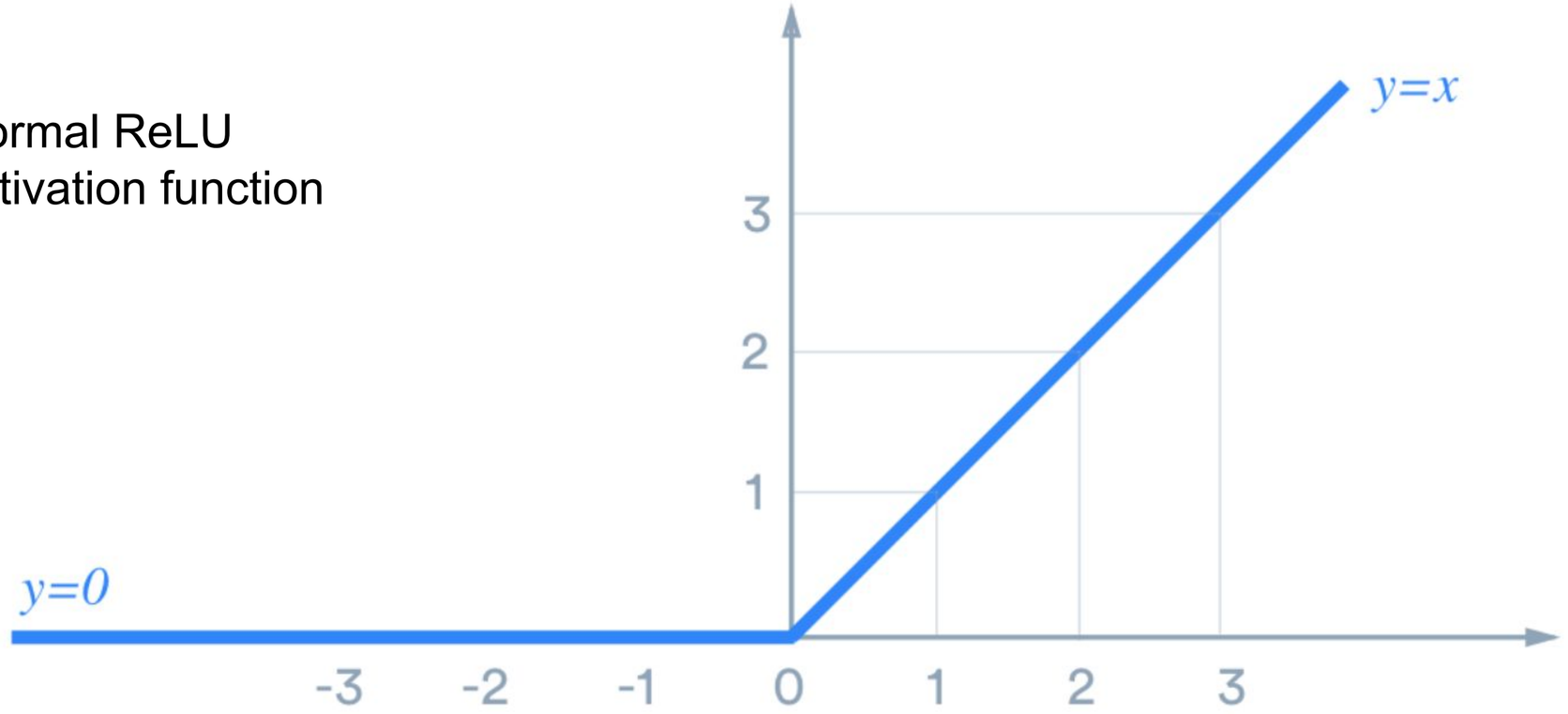


ReLU
Rectified Linear Unit

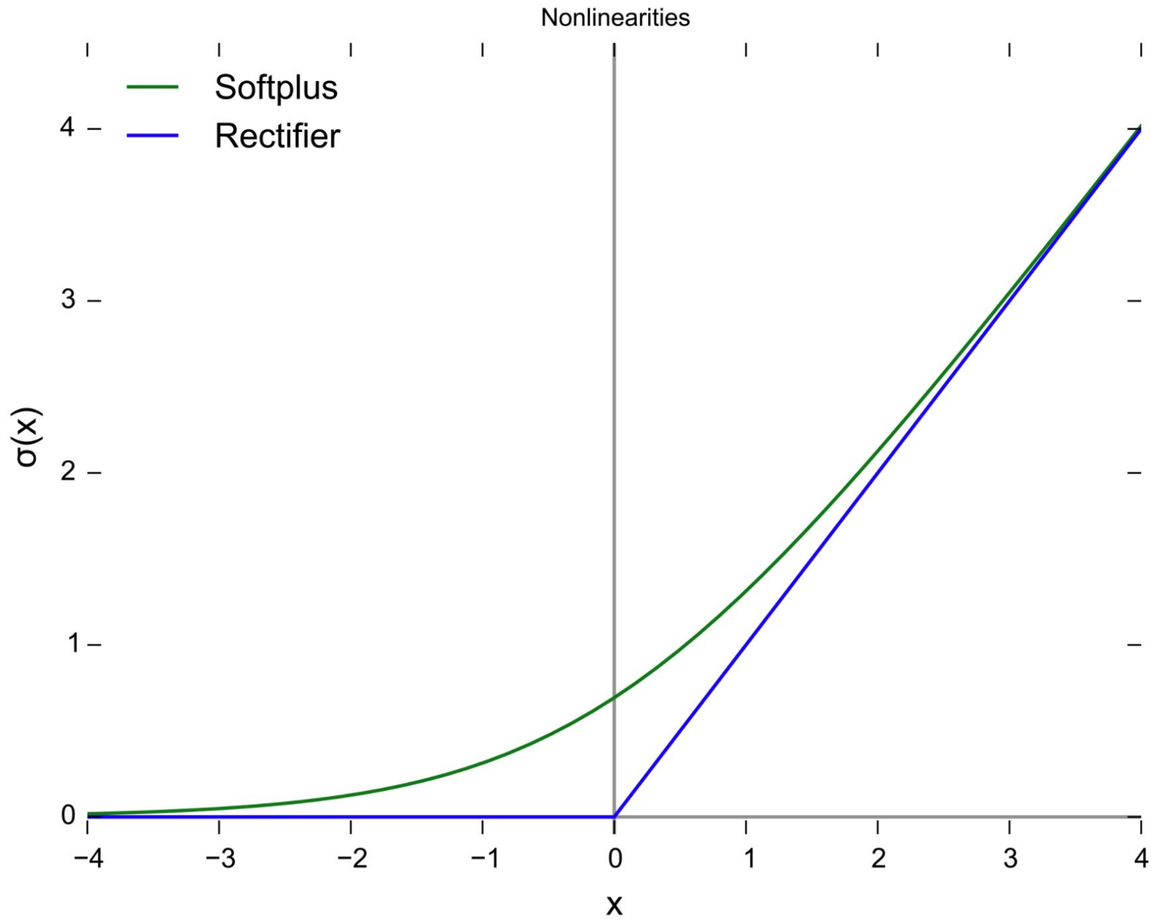
$$f(x) = max(0, x)$$

There are many different ReLU variants

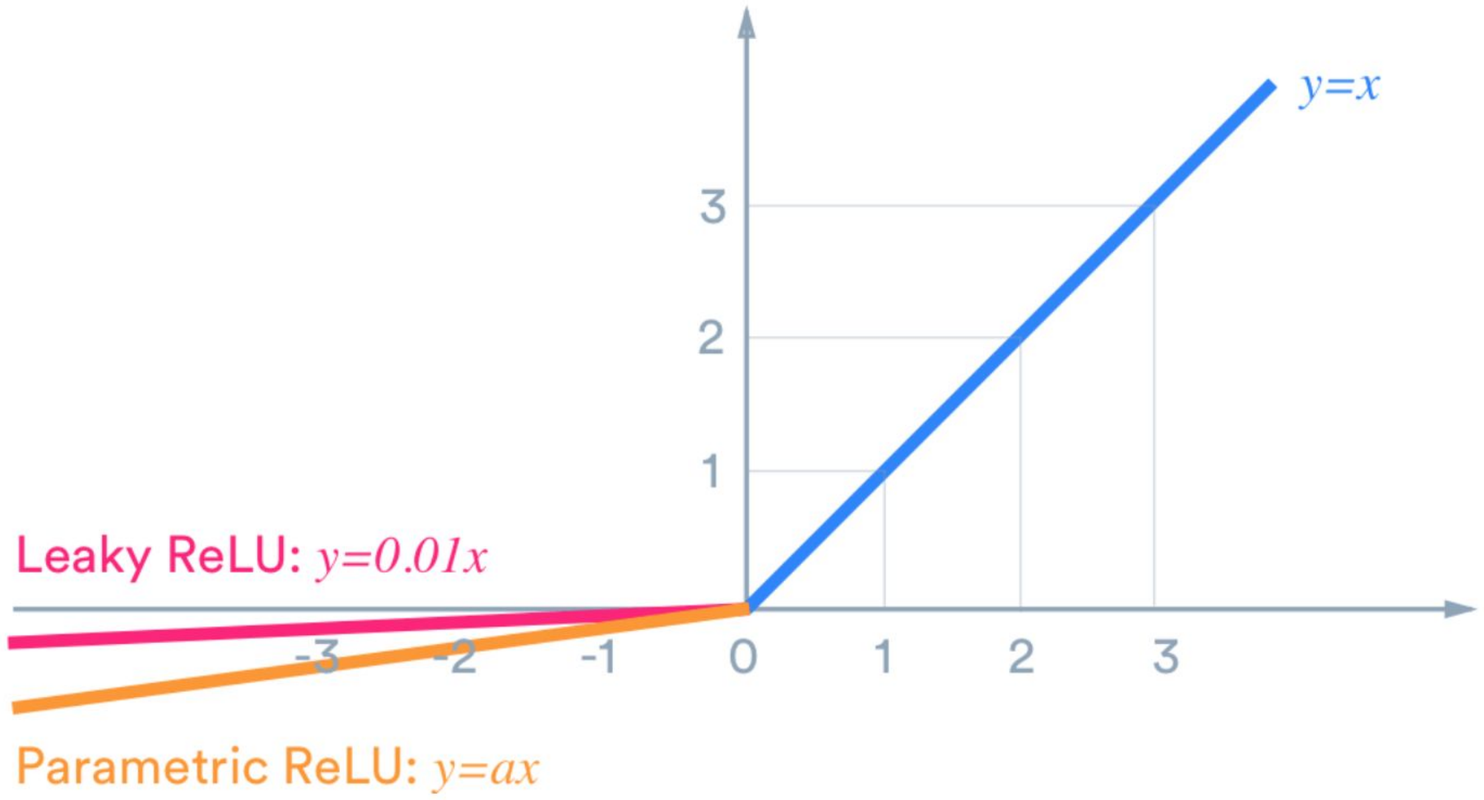
Normal ReLU
activation function



There are many different ReLU variants

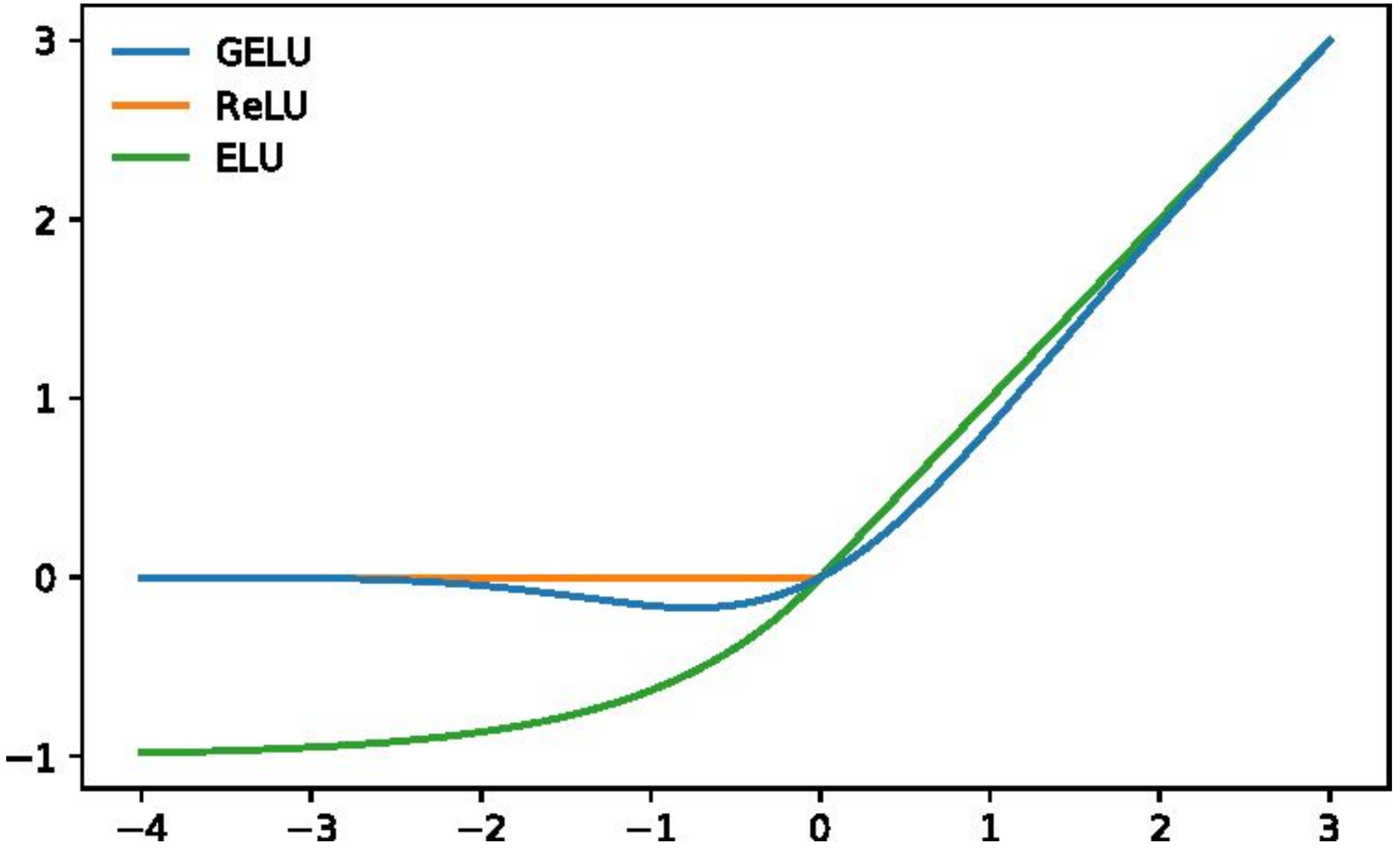


There are many different ReLU variants

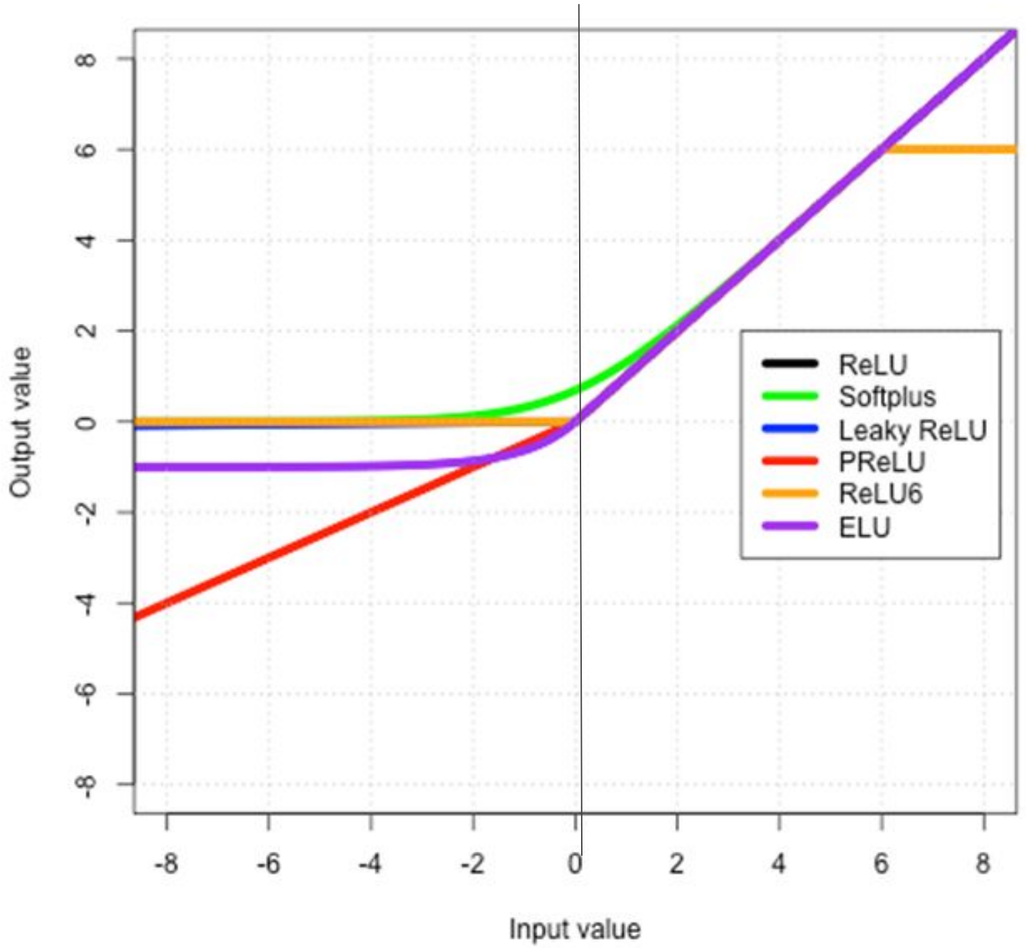


Leaky ReLU

There are many different ReLU variants



There are many different ReLU variants



Three common failure modes for gradient descent

#1 Gradients can vanish — Problem

Each additional layer
can successively reduce
signal vs. noise — Insight

**Using ReLu instead of
sigmoid/tanh can help** — Solution

Three common failure modes for gradient descent

Gradients can vanish	**#2 Gradients can explode**	Problem
Each additional layer can successively reduce signal vs. noise	Learning rates are important here	Insight
Using ReLu instead of sigmoid/tanh can help	**Batch normalization (useful knob) can help**	Solution

Three common failure modes for gradient descent

Gradients can vanish	**Gradients can explode**	**#3 ReLu layers can die**	Problem
Each additional layer can successively reduce signal vs. noise	Learning rates are important here	Monitor fraction of zero weights in TensorBoard	Insight
Using ReLu instead of sigmoid/tanh can help	**Batch normalization (useful knob) can help**	**Lower your learning rates**	Solution

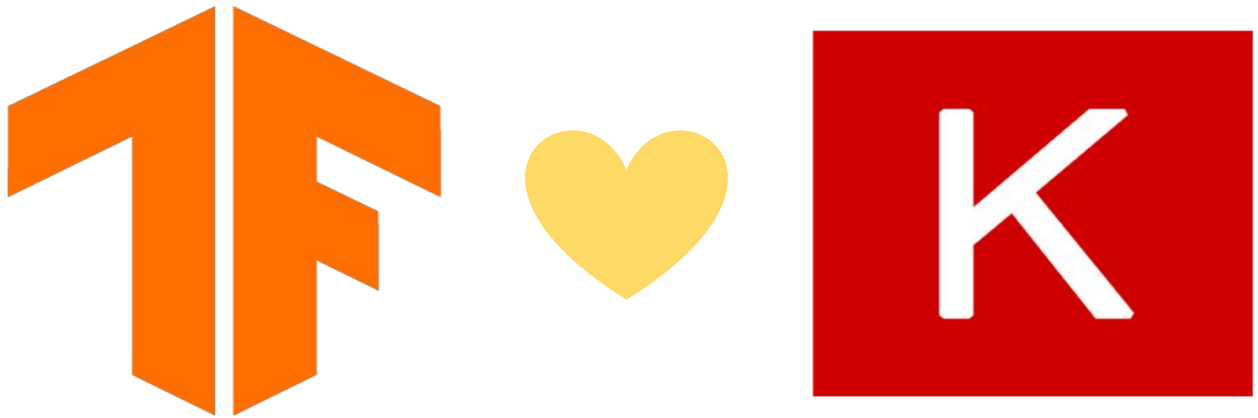
Agenda

Activation Functions

Neural Networks with TF 2 and Keras

Regularization

Keras is built-in to TF 2.x



Stacking layers with Keras Sequential model

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense


model = Sequential([
    Input(shape=(64,))
    Dense(units=32, activation="relu", name="hidden1"),
    Dense(units=8, activation="relu", name="hidden2"),
    Dense(units=1, activation="linear", name="output")
])
```

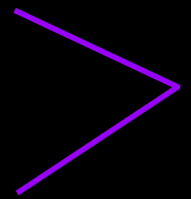
The batch size is omitted. Here the model expects batches of vectors with 64 components.

The Keras sequential model stacks layers on the top of each other.

```
%tensorflow_version 2.x
import tensorflow as tf
from tensorflow import keras
(x_train, y_train), (x_test, y_test) = keras.datasets.mnist.load_data()



# Define your model
model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(10, activation='softmax')
])
```

```
%tensorflow_version 2.x
import tensorflow as tf
from tensorflow import keras
(x_train, y_train), (x_test, y_test) = keras.datasets.mnist.load_data()
```

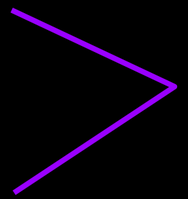
```
# Define your model
model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(10, activation='softmax')
])
```



A linear model (a single Dense layer) aka multiclass logistic regression

```
%tensorflow_version 2.x
import tensorflow as tf
from tensorflow import keras
(x_train, y_train), (x_test, y_test) = keras.datasets.mnist.load_data()
```

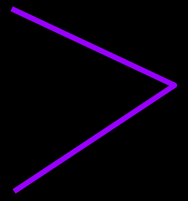
```
# Define your model
model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(10, activation='softmax')
])
```



A neural network with one hidden layer

```
%tensorflow_version 2.x
import tensorflow as tf
from tensorflow import keras
(x_train, y_train), (x_test, y_test) = keras.datasets.mnist.load_data()
```

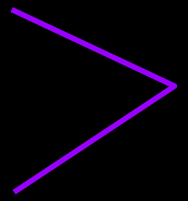
```
# Define your model
model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(10, activation='softmax')
])
```



A neural network with multiple hidden layers (a deep neural network)

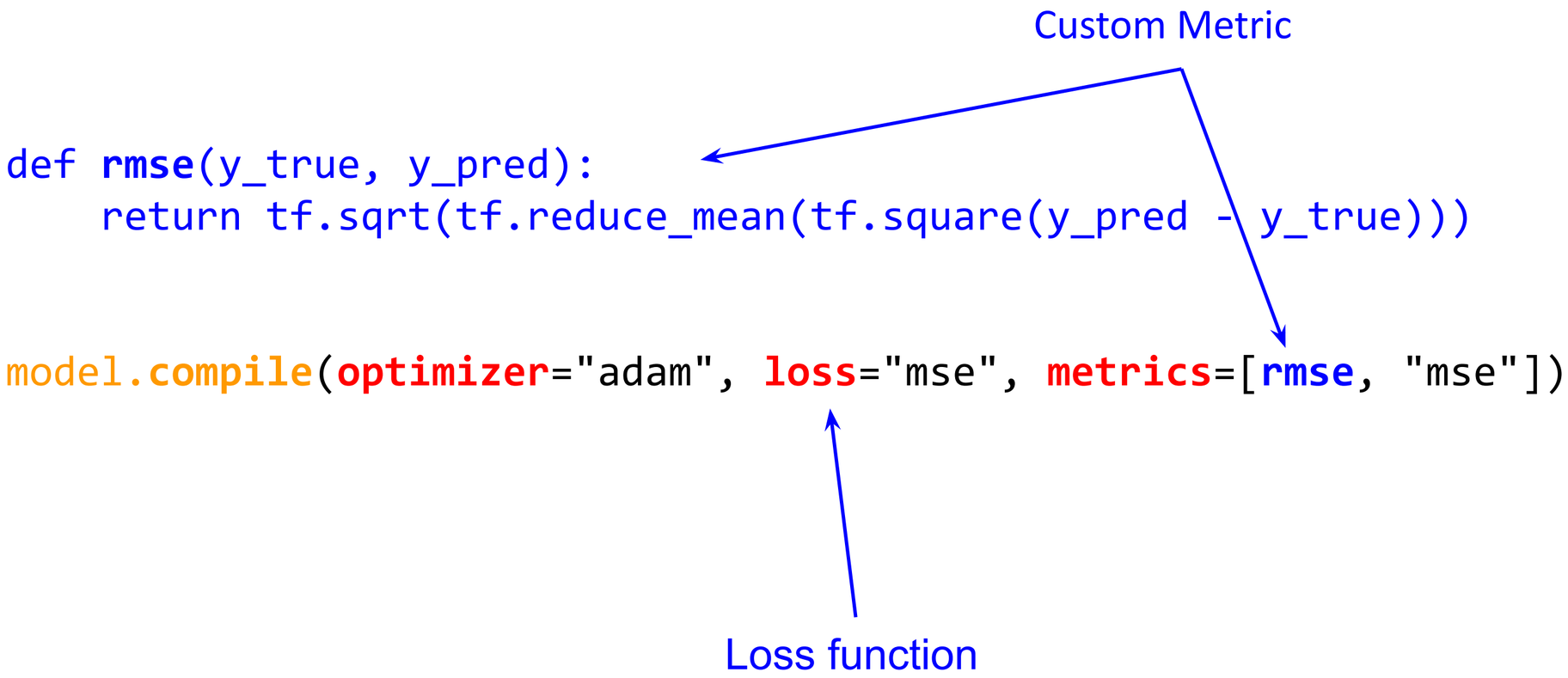
```
%tensorflow_version 2.x
import tensorflow as tf
from tensorflow import keras
(x_train, y_train), (x_test, y_test) = keras.datasets.mnist.load_data()
```

```
# Define your model
model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(10, activation='softmax')
])
```



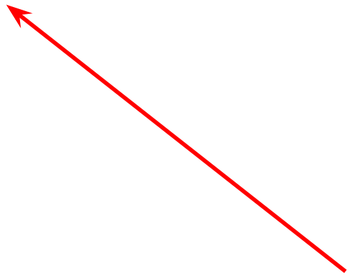
A deeper neural network

Compiling a Keras model



Compiling a Keras model

```
def rmse(y_true, y_pred):
    return tf.sqrt(tf.reduce_mean(tf.square(y_pred - y_true)))


model.compile(optimizer="adam", loss="mse", metrics=[rmse, "mse"])
```



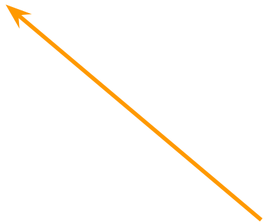
Optimizer

Compiling a Keras model

```
def rmse(y_true, y_pred):
    return tf.sqrt(tf.reduce_mean(tf.square(y_pred - y_true)))


model.compile(optimizer="adam", loss="mse", metrics=[rmse, "mse"])
```

Training a Keras model

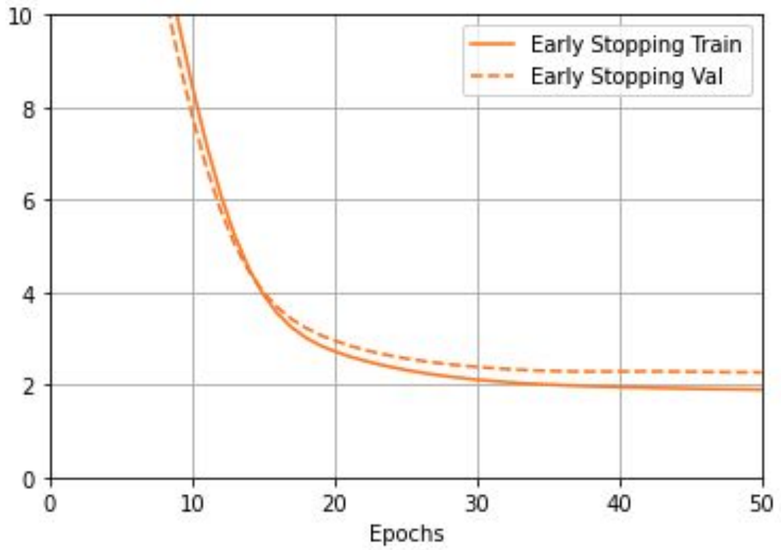
```
from tensorflow.keras.callbacks import TensorBoard

steps_per_epoch = NUM_TRAIN_EXAMPLES // (TRAIN_BATCH_SIZE * NUM_EVALS)

history = model.fit(
    x=trainds,
    steps_per_epoch=steps_per_epoch,
    epochs=NUM_EVALS,
    validation_data=evalds,
    callbacks=[TensorBoard(LOGDIR)]
)
```



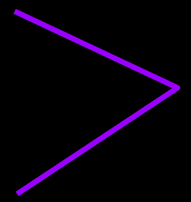
This is a trick so that we have control on the total number of examples the model trains on (NUM_TRAIN_EXAMPLES) and the total number of evaluation we want to have during training (NUM_EVALS).

Training a Keras model

```
from tensorflow.keras.callbacks import TensorBoard

steps_per_epoch = NUM_TRAIN_EXAMPLES // (TRAIN_BATCH_SIZE * NUM_EVALS)

history = model.fit(
    x=trainds,
    steps_per_epoch=steps_per_epoch,
    epochs=NUM_EVALS,
    validation_data=evalds,
    callbacks=[TensorBoard(LOGDIR)]
)
```



```
%tensorflow_version 2.x
import tensorflow as tf
from tensorflow import keras
(x_train, y_train), (x_test, y_test) = keras.datasets.mnist.load_data()

# Define your model
model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dense(10, activation='softmax')
])

# Configure and train
model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])
model.fit(x_train, y_train, epochs=5)
model.evaluate(x_test, y_test)
```



A deeper neural network

Once trained, the model can be used for prediction

```
predictions = model.predict(input_samples, steps=1)
```

returns a Numpy array of predictions

With the `predict()` method you can pass

- a Dataset instance
- Numpy array
- a Tensorflow tensor, or list of tensors
- a generator of input samples

`steps` determines the total number of steps before declaring the prediction round finished. Here, since we have just one example, `steps=1`.

To serve our model for others to use, we export the model file and deploy the model as a service.

SavedModel is the universal serialization format for Tensorflow models

```
OUTPUT_DIR = "./export/savedmodel"
shutil.rmtree(OUTPUT_DIR, ignore_errors=True)

EXPORT_PATH = os.path.join(OUTPUT_DIR,
datetime.datetime.now().strftime("%Y%m%d%H%M%S"))

tf.saved_model.save(model, EXPORT_PATH)
```

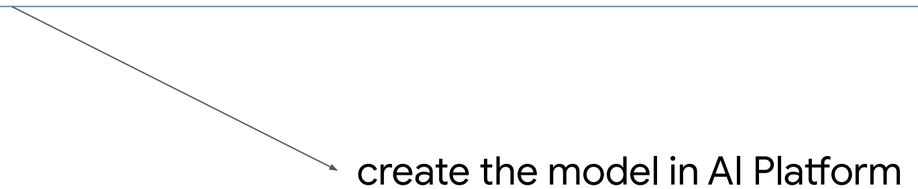
the directory in which to write the SavedModel

a trackable object such as a trained keras model

exports a model object to a SavedModel format

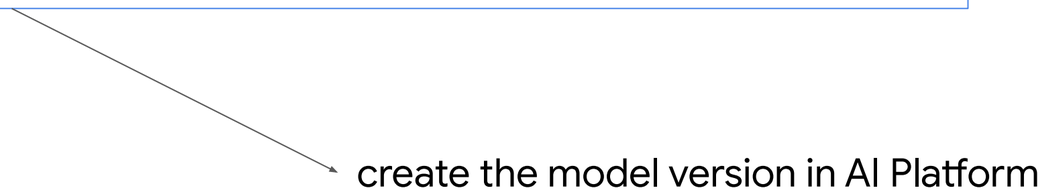
Create a model object in Cloud AI Platform

```
MODEL_NAME=propertyprice
VERSION_NAME=dnn

if [[ $(gcloud ai-platform models list --format='value(name)' | grep $MODEL_NAME) ]];
then
    echo "$MODEL_NAME already exists"
else
    echo "Creating $MODEL_NAME"
    gcloud ai-platform models create --regions=$REGION $MODEL_NAME
fi

...
```



Create a version of the model in Cloud AI Platform

```
MODEL_NAME=propertyprice
VERSION_NAME=dnn


...


if [[ $(gcloud ai-platform versions list --model $MODEL_NAME --format='value(name)' |
grep $VERSION_NAME) ]]; then
    echo "Deleting already existing $MODEL_NAME:$VERSION_NAME ... "
    echo yes | gcloud ai-platform versions delete --model=$MODEL_NAME $VERSION_NAME
    echo "Please run this cell again if you don't see a Creating message ... "
    sleep 2
fi
```



Deploy SavedModel using `gcloud ai-platform`

```
gcloud ai-platform versions create \
        --model=$MODEL_NAME $VERSION_NAME \
        --framework=tensorflow \
        --python-version=3.5 \
        --runtime-version=2.1 \
        --origin=$EXPORT_PATH \
        --staging-bucket=gs://$BUCKET
```

specify the model name and version

`EXPORT_PATH` denotes location of SavedModel directory

Make predictions using `gcloud ai-platform`

```
input.json = {"sq_footage": 3140,
              "type": 'house'}
```

```
gcloud ai-platform predict \
    --model propertyprice \
    --version dnn \
    --json-instances input.json
```

specify the name and version of the deployed model

json file for prediction

Lab

Keras Sequential API

In this lab, you will build a DNN model to predict the fare amount using Keras sequential API

..deepdive2//introduction_to_tensorflow/labs/3_keras_sequential_api.ipynb

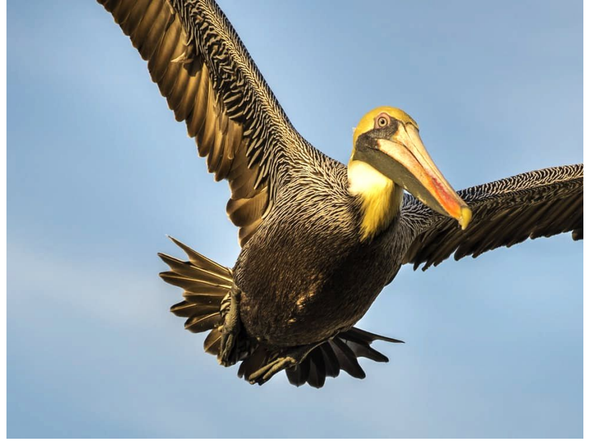
Google Cloud

DNNs with the Keras Functional API

Seagulls
can fly.

Pigeons
can fly.

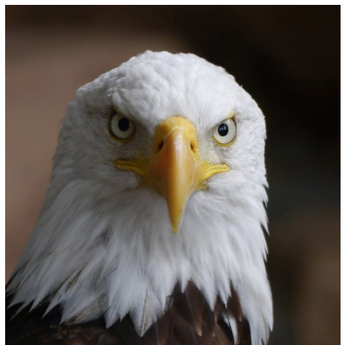










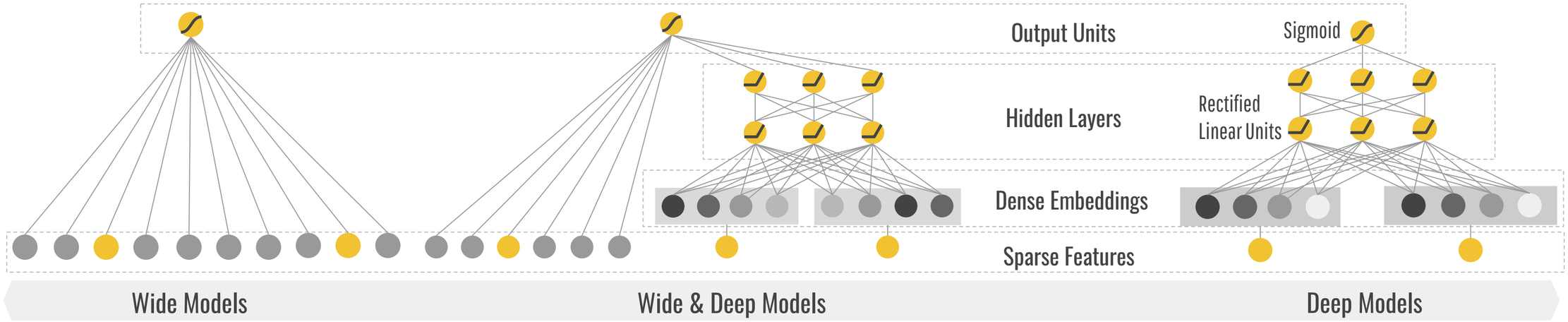



Animals with wings can fly.

Penguins...

Usings Wide and Deep learning

“ Combine the power of memorization and generalization on one unified machine learning model. ”



Memorization + Generalization

- Memorization: “Seagulls can fly.” “Pigeons can fly.”
- Generalization: “**Animals with wings** can fly.”
- Generalization + memorizing exceptions: “Animals with wings can fly, but penguins cannot fly.”

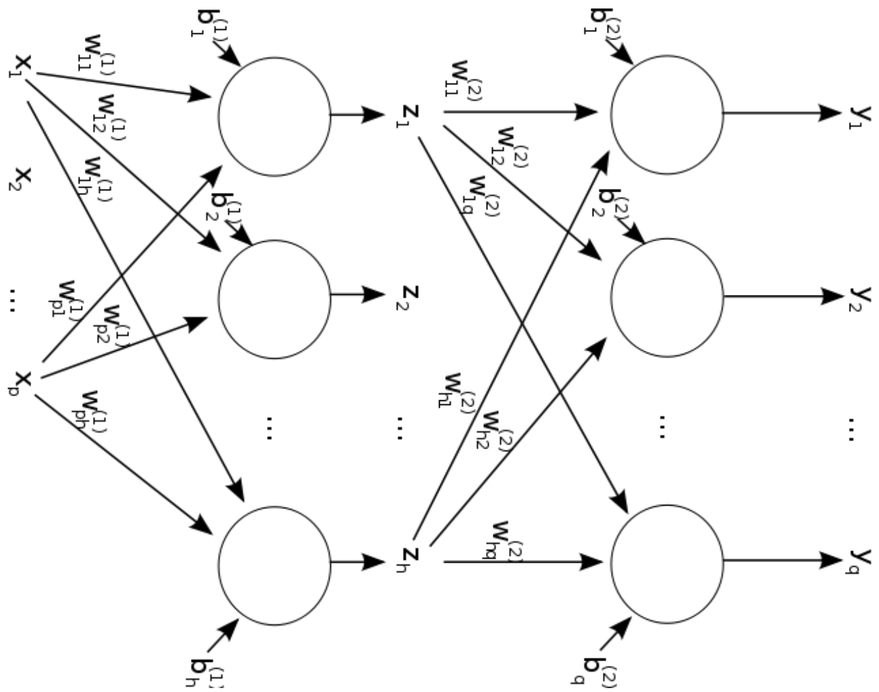
Linear models are good for sparse, independent features

```
[ 0.  1.  0.  0.  0.  0.  0.  0.  0.  0.]
[ 0.  0.  0.  0.  0.  0.  0.  1.  0.  0.]
[ 0.  0.  0.  0.  0.  0.  0.  0.  0.  1.]
[ 0.  1.  0.  0.  0.  0.  0.  0.  0.  0.]
[ 0.  0.  0.  0.  0.  0.  0.  1.  0.  0.]
[ 0.  0.  0.  1.  0.  0.  0.  0.  0.  0.]
[ 0.  0.  0.  0.  0.  0.  1.  0.  0.  0.]
[ 0.  1.  0.  0.  0.  0.  0.  0.  0.  0.]
[ 0.  0.  0.  0.  1.  0.  0.  0.  0.  0.]
[ 0.  1.  0.  0.  0.  0.  0.  0.  0.  0.]
```

DNNs are good for dense, highly correlated features

pixel_values (

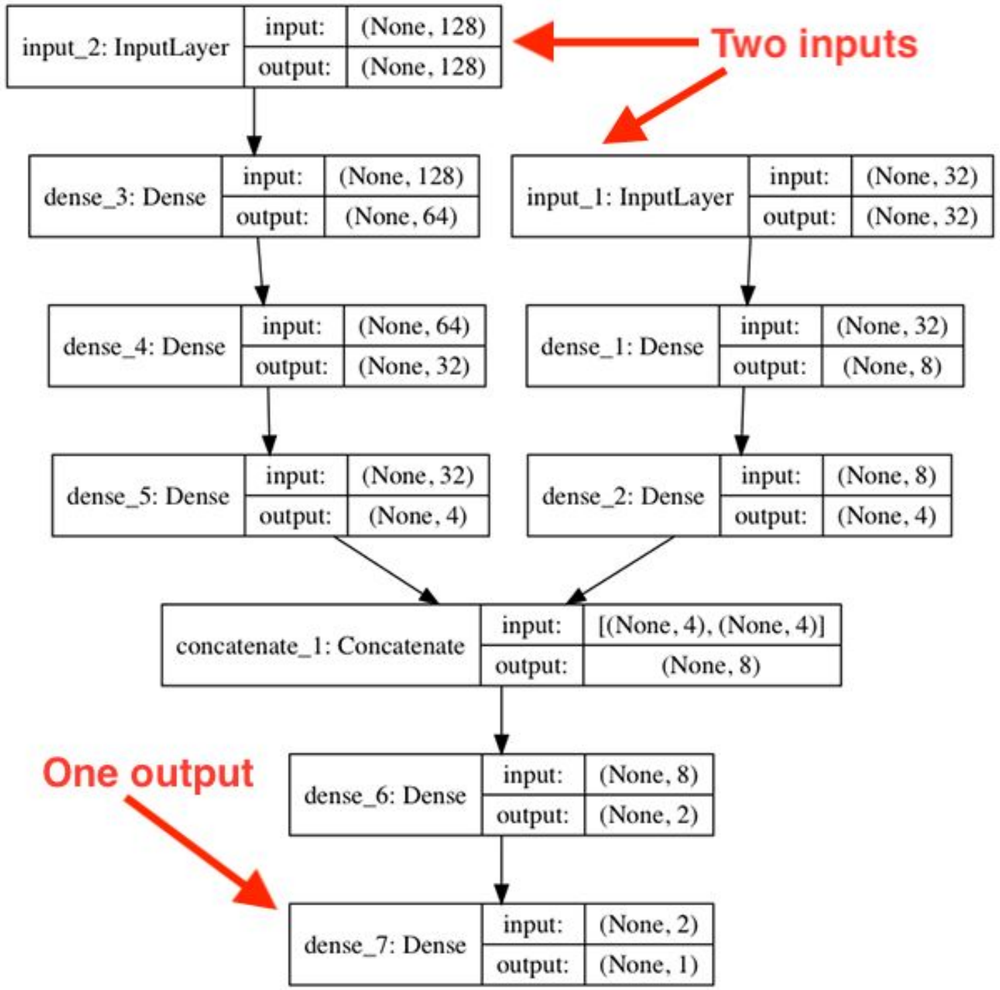
)



1024^2 input nodes

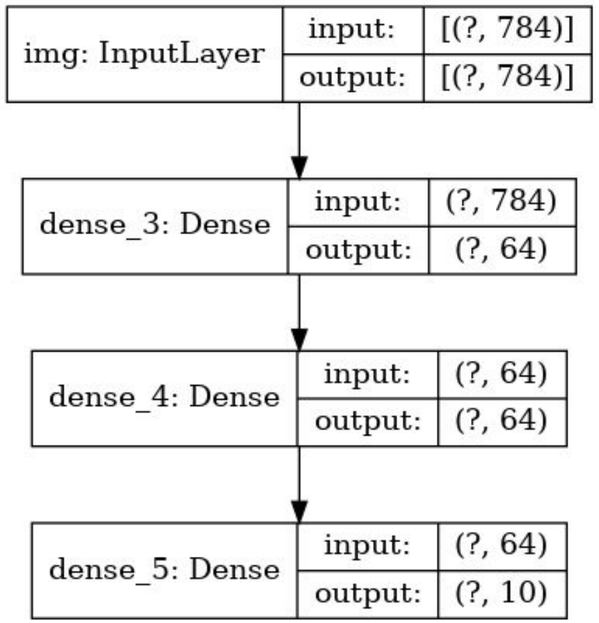
10 hidden nodes -> 10 image features

Wide and deep networks using Keras Functional API



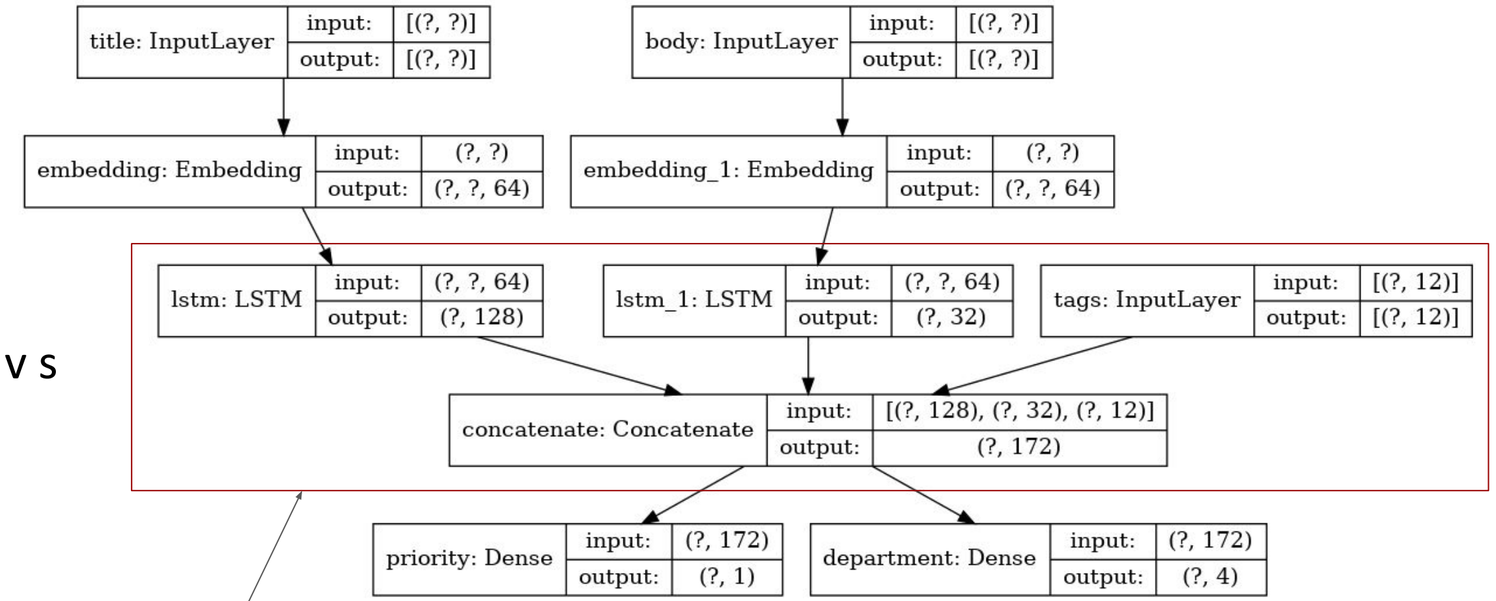
Functional API is more flexible than Sequential API

Sequential model in Keras



v s

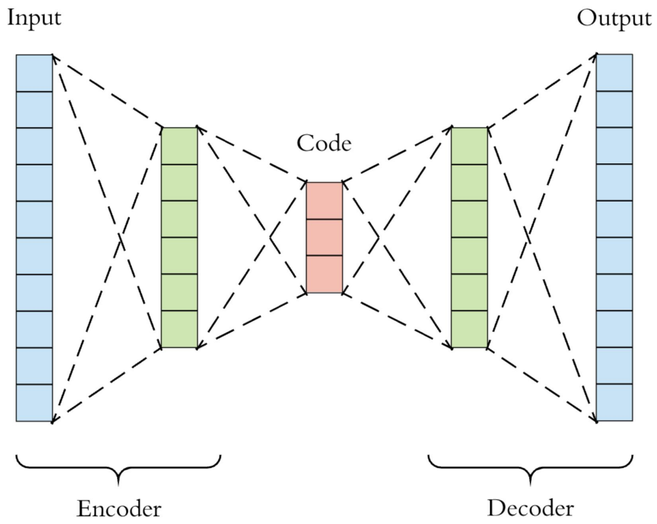
Functional model in Keras



Multiple inputs / shared layer

Models are created by specifying their inputs and outputs in a graph of layers

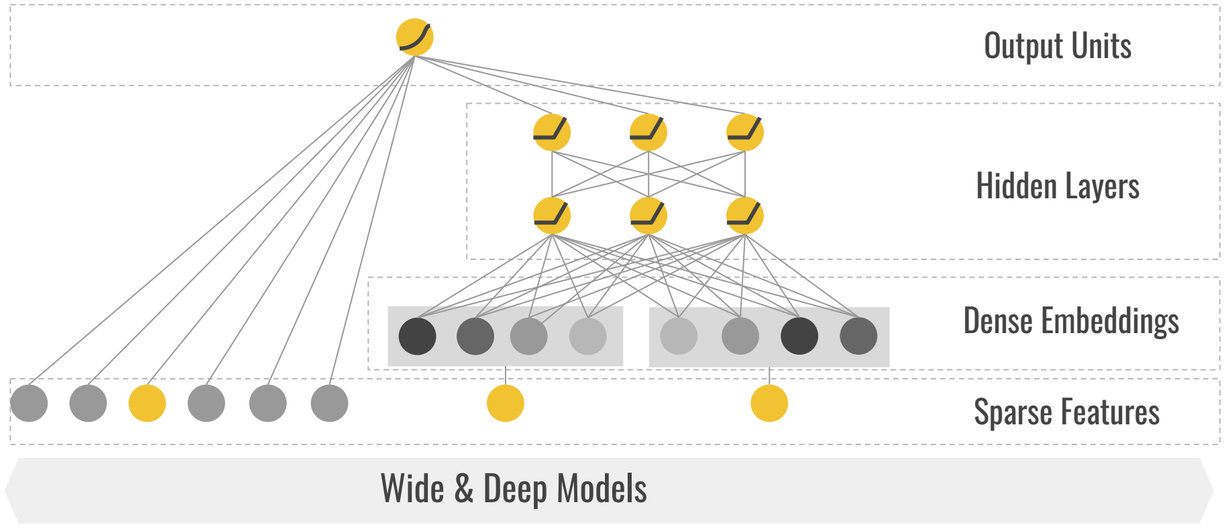
```
encoder_input = keras.Input(shape=(28, 28, 1), name='img')
x = layers.Dense(16, activation='relu')(encoder_input)
x = layers.Dense(10, activation='relu')(x)
x = layers.Dense(5, activation='relu')(x)
encoder_output = layers.Dense(3, activation='relu')(x)

encoder = keras.Model(encoder_input, encoder_output, name='encoder')
```




```
x = layers.Dense(5, activation='relu')(encoder_output)
x = layers.Dense(10, activation='relu')(x)
x = layers.Dense(16, activation='relu')(x)
decoder_output = layers.Dense(28, activation='linear')(x)

autoencoder = keras.Model(encoder_input, decoder_output, name='autoencoder')
```

Creating a Wide and Deep model in Keras

```
INPUT_COLS = [
    'pickup_longitude',
    'pickup_latitude',
    'dropoff_longitude',
    'dropoff_latitude',
    'passenger_count'
]
```

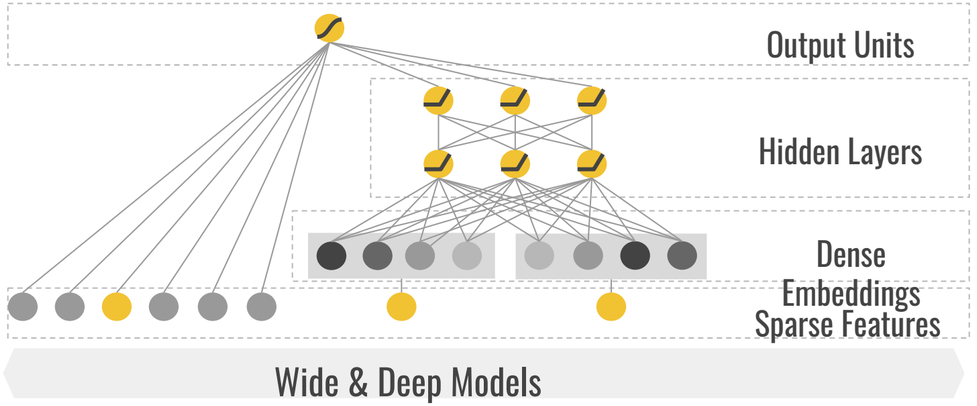



```
# Prepare input feature columns
inputs = {colname : layers.Input(name=colname, shape=(), dtype='float32')
          for colname in INPUT_COLS
}
```

...

Creating a Wide and Deep model in Keras

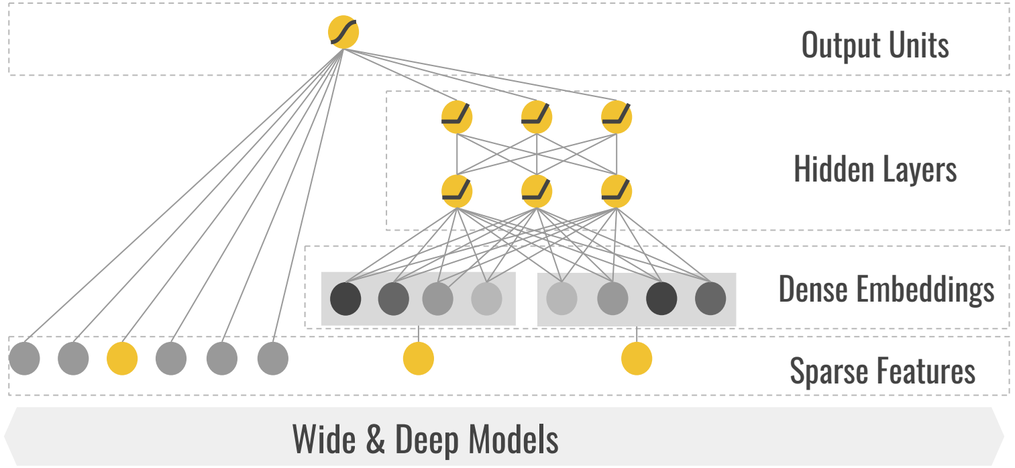
```
# Create deep columns
deep_columns = [
    # Embedding_column to "group" together ...
    fc.embedding_column(fc_crossed_pd_pair, 10),

    # Numeric columns
    fc.numeric_column("pickup_latitude"),
    fc.numeric_column("pickup_longitude"),
    fc.numeric_column("dropoff_longitude"),
    fc.numeric_column("dropoff_latitude")]

# Create the deep part of model
deep_inputs = layers.DenseFeatures
      (deep_columns, name='deep_inputs')(inputs)
x = layers.Dense(30, activation='relu')(deep_inputs)
x = layers.Dense(20, activation='relu')(x)

deep_output = layers.Dense(10, activation='relu')(x)
```



created in the previous slide

Creating a Wide and Deep model in Keras

```
# Create wide columns

wide_columns = [
    # One-hot encoded feature crosses
    fc.indicator_column(fc_crossed_dloc),
    fc.indicator_column(fc_crossed_ploc),
    fc.indicator_column(fc_crossed_pd_pair)
]
```

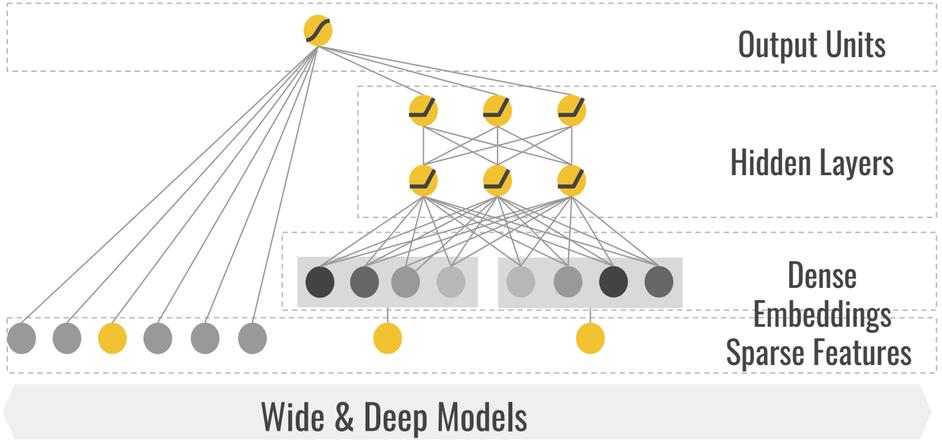



```
# Create the wide part of model

wide = layers.DenseFeatures(wide_columns, name='wide_inputs')(inputs)
```

created in the previous slide

Creating a Wide and Deep model in Keras

```
# Combine outputs
combined = concatenate(inputs=[deep, wide],
                       name='combined')
output = layers.Dense(1,
                      activation=None,
                      name='prediction')(combined)
```




```
# Finalize model
model = keras.Model(inputs=list(inputs.values()),
                    outputs=output,
                    name='wide_and_deep')
model.compile(optimizer="adam",
              loss="mse",
              metrics=[rmse, "mse"])
```

To finalize the model, specify the inputs and outputs

Strengths and weaknesses of the Functional API

Strengths

- less verbose than using keras.Model subclasses
- validates your model while you're defining it
- your model is plottable and inspectable
- your model can be serialized or cloned

Strengths and weaknesses of the Functional API

Strengths

- less verbose than using keras.Model subclasses
- validates your model while you're defining it
- your model is plottable and inspectable
- your model can be serialized or cloned

Weaknesses

- doesn't support dynamic architectures
- sometimes you have to write from scratch and you need to build subclasses, e.g. custom training or inference layers

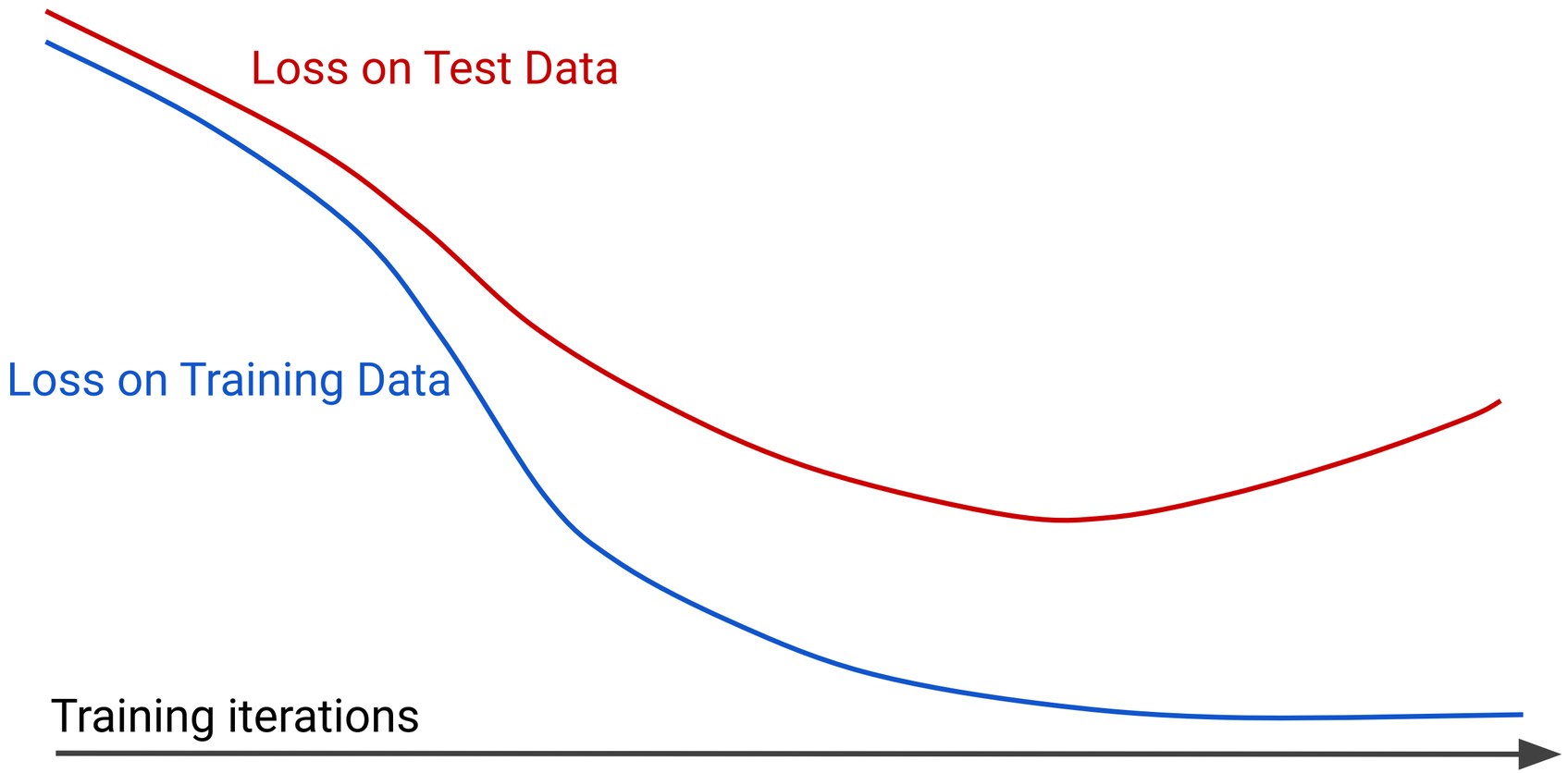
Agenda

Activation Functions

Neural Networks with TF 2 and Keras

Regularization

What's happening here? How can we address this?



Why does it happen?

https://goo.gl/ofiHCT

Is the model behavior surprising? What's the issue?

Try removing cross-product features. Does performance improve?

The simpler the better

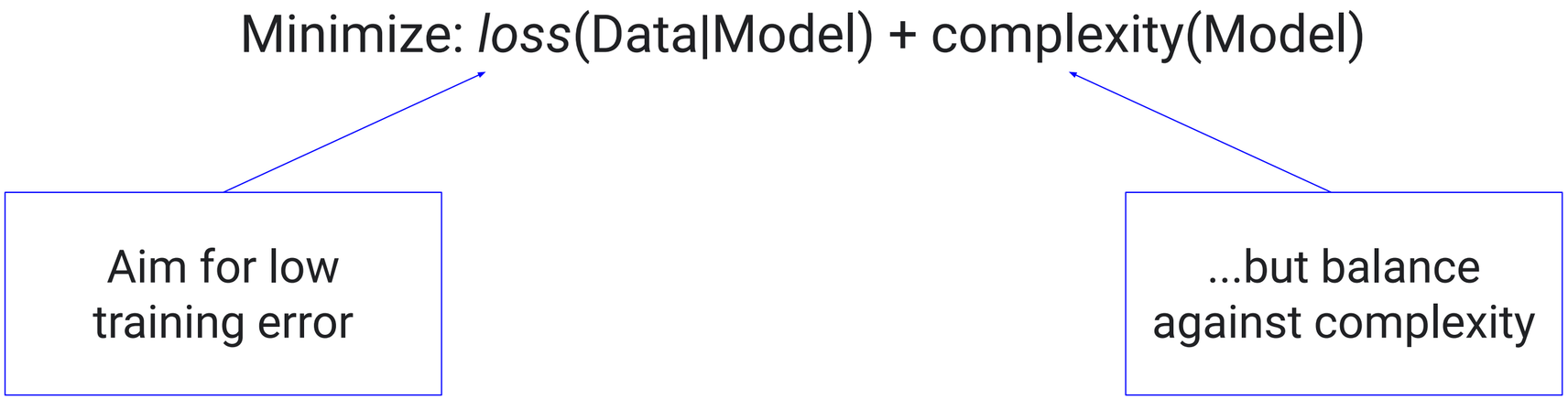
Don’t cook with every spice in the spice rack!

When presented with competing hypothetical answers to a problem, one should select the one that makes the fewest assumptions.

The idea is attributed to William of Ockham (c. 1287–1347).

source: https://en.wikipedia.org/wiki/Occam%27s_razor

Factor in model complexity when calculating error



Optimal model complexity is data-dependent, so requires hyperparameter tuning.

Regularization is a major field of ML research

Early Stopping

Parameter Norm Penalties

- L1 regularization
- L2 regularization
- Max-norm regularization

Dataset Augmentation

Noise Robustness

Sparse Representations

. . .

We'll look into these methods.

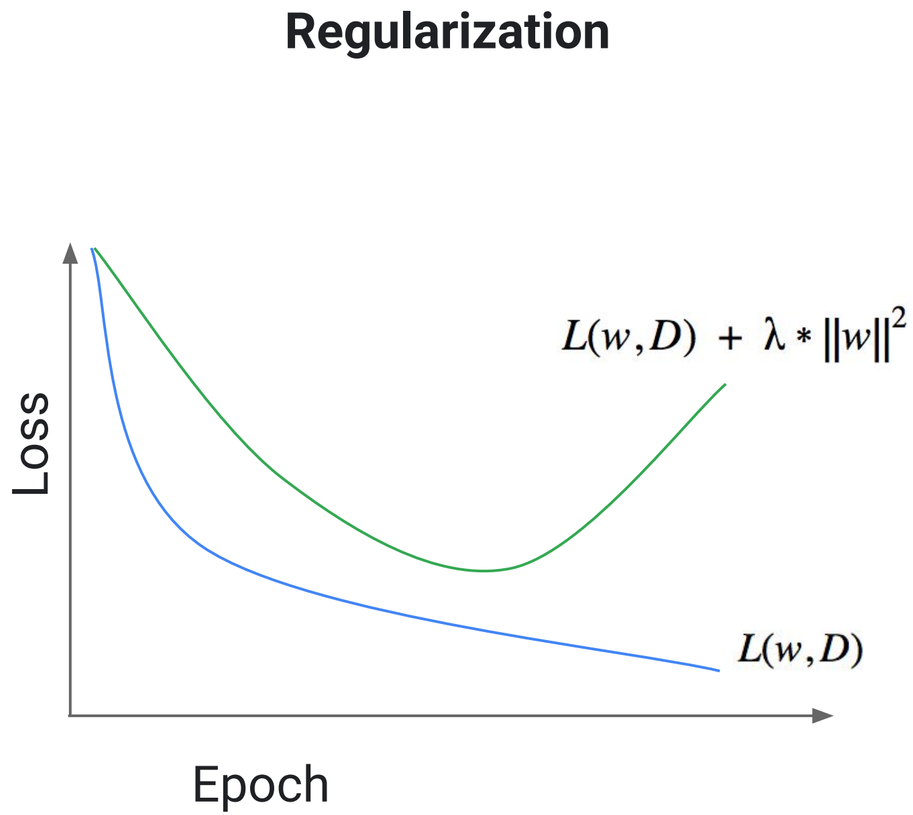
Zeroing out coefficients can help with performance, especially with large models and sparse inputs

Action	Impact
Fewer coefficients to store/load.	Reduce memory, model size.
Fewer multiplications needed.	Increase prediction speed.

$$L(w, D) + \lambda \sum^{n} |w|$$

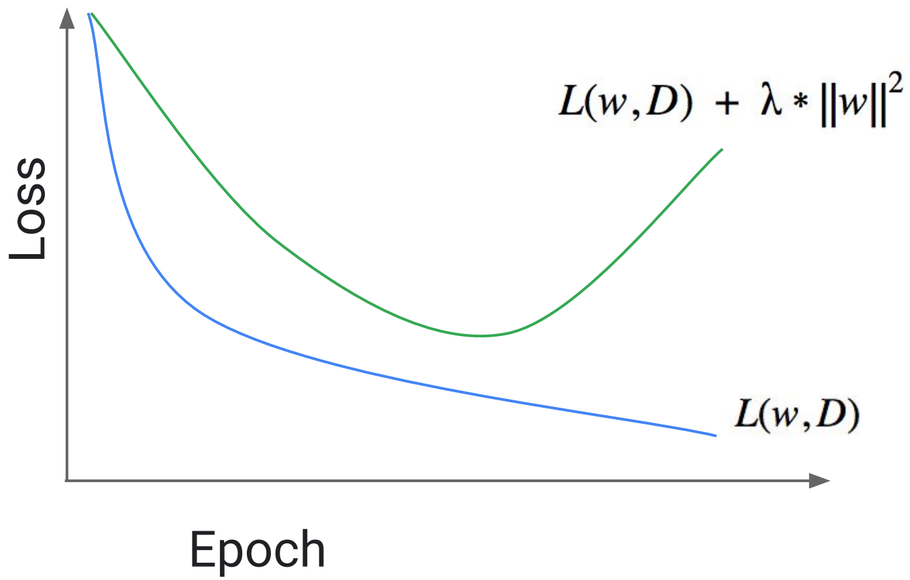
L2 regularization only makes weights small, not zero.

Often we do both regularization and early stopping to counteract overfitting

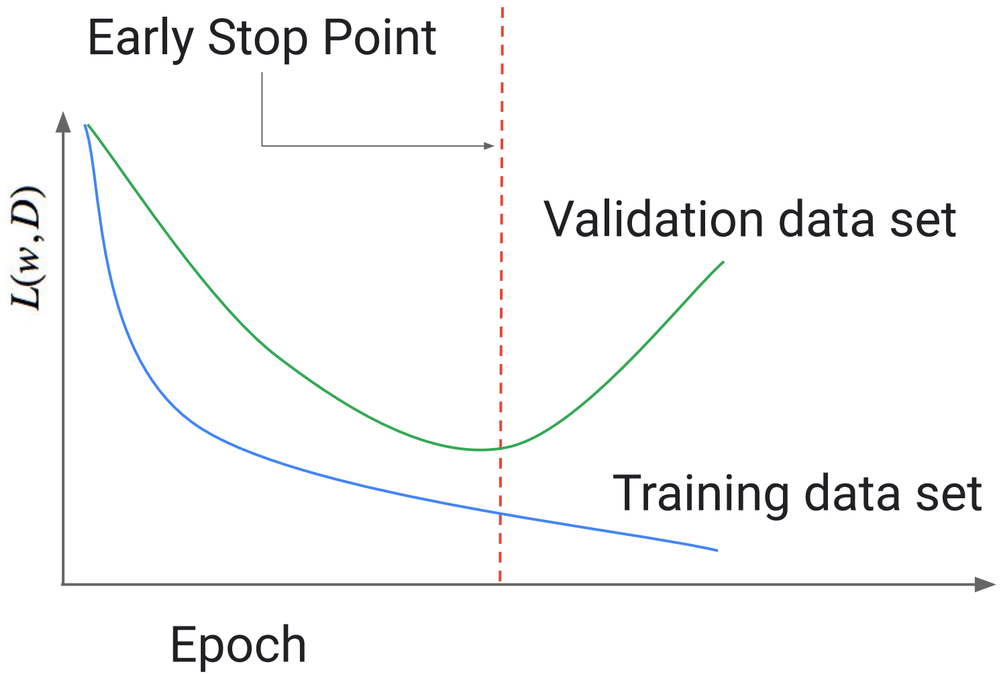


Often we do both regularization and early stopping to counteract overfitting

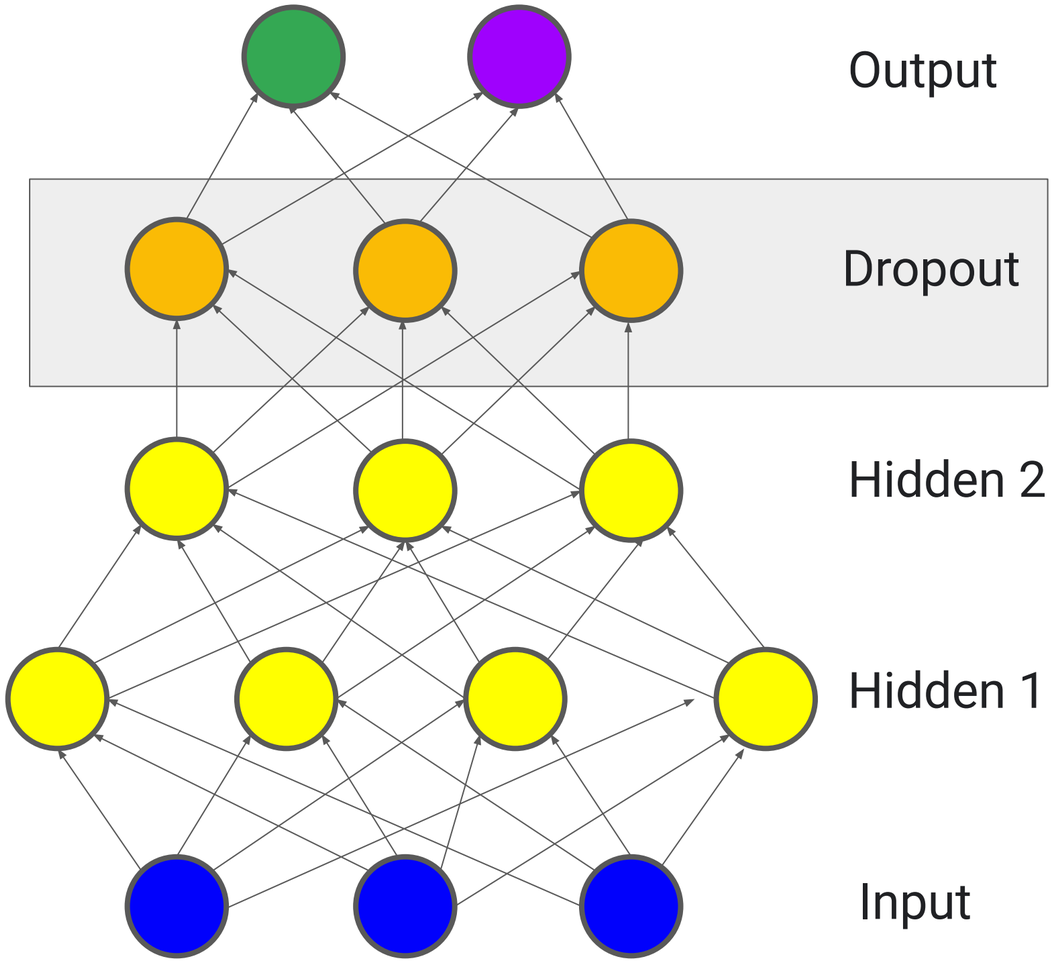
Regularization



Early stopping

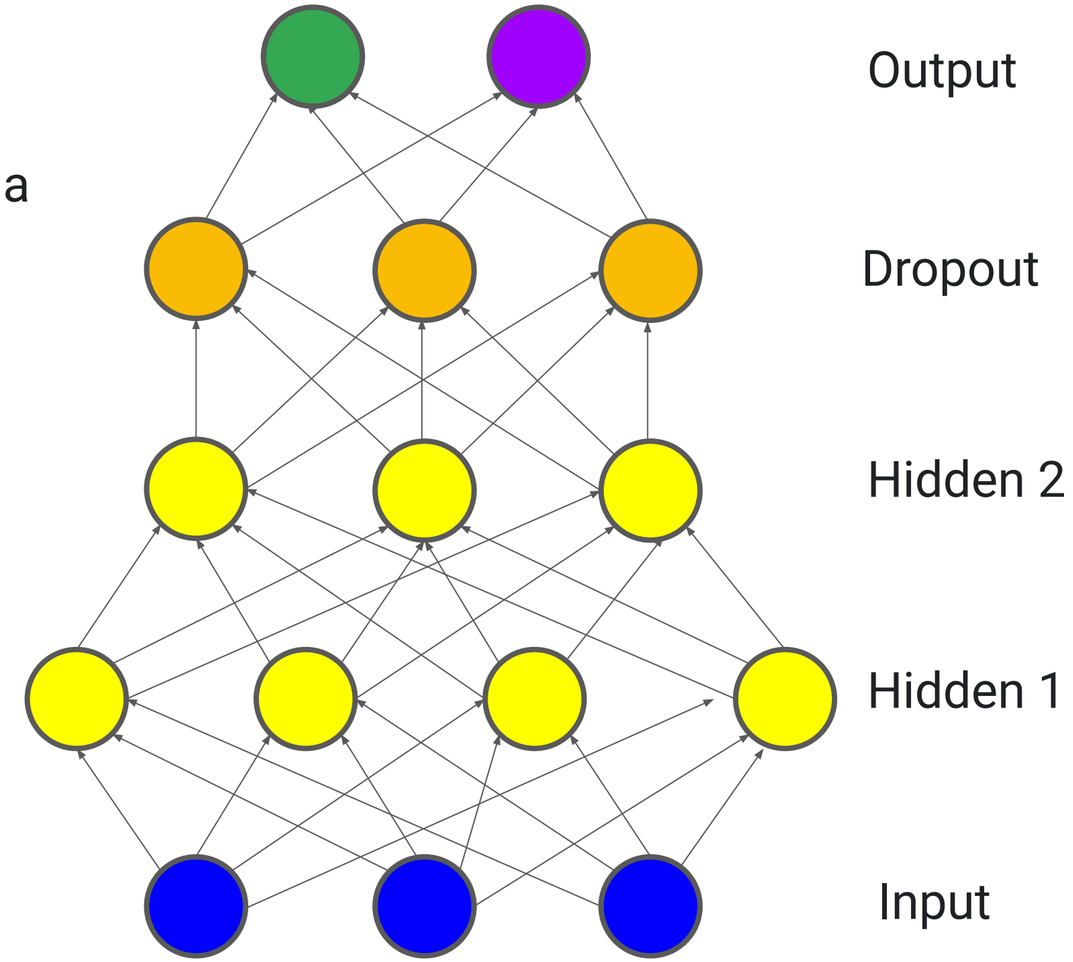


Dropout layers are a form of regularization



Dropout layers are a form of regularization

Dropout works by randomly “dropping out” unit activations in a network for a single gradient step.

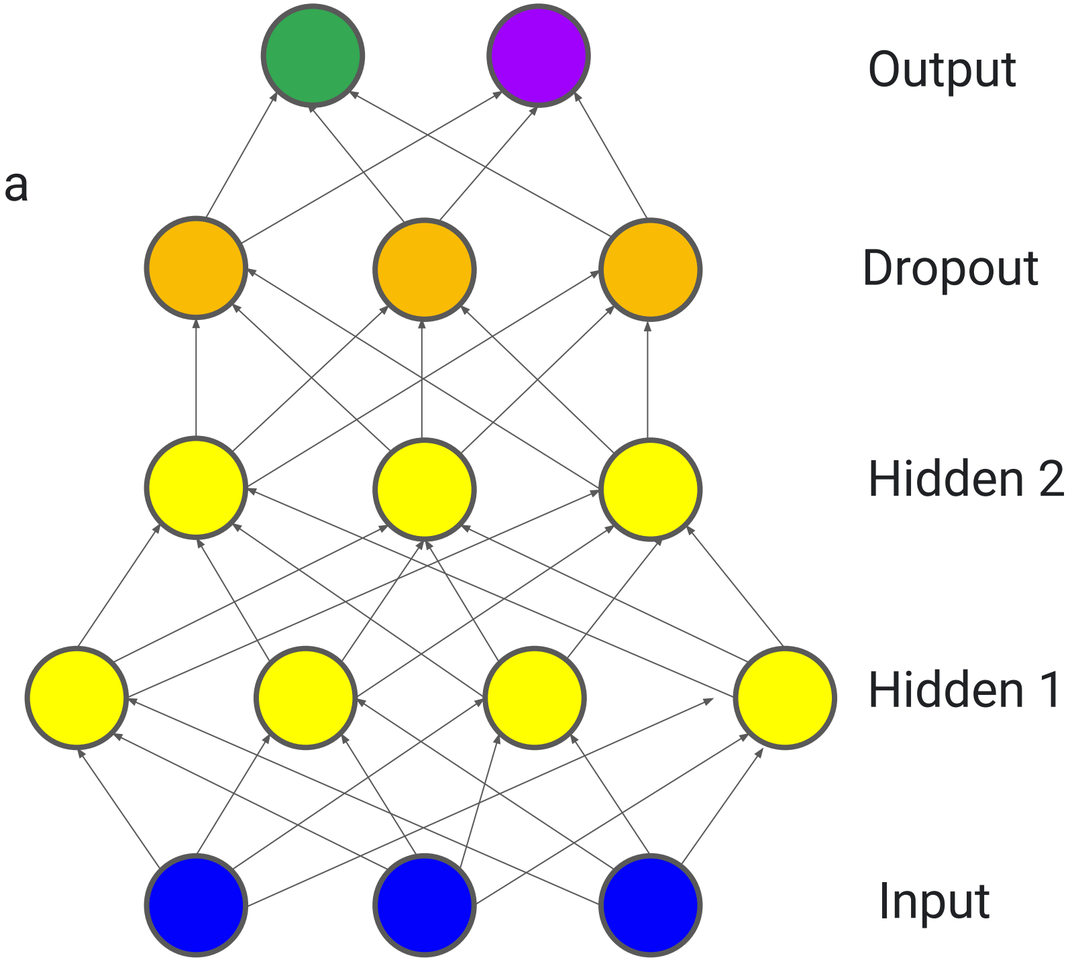


Dropout layers are a form of regularization

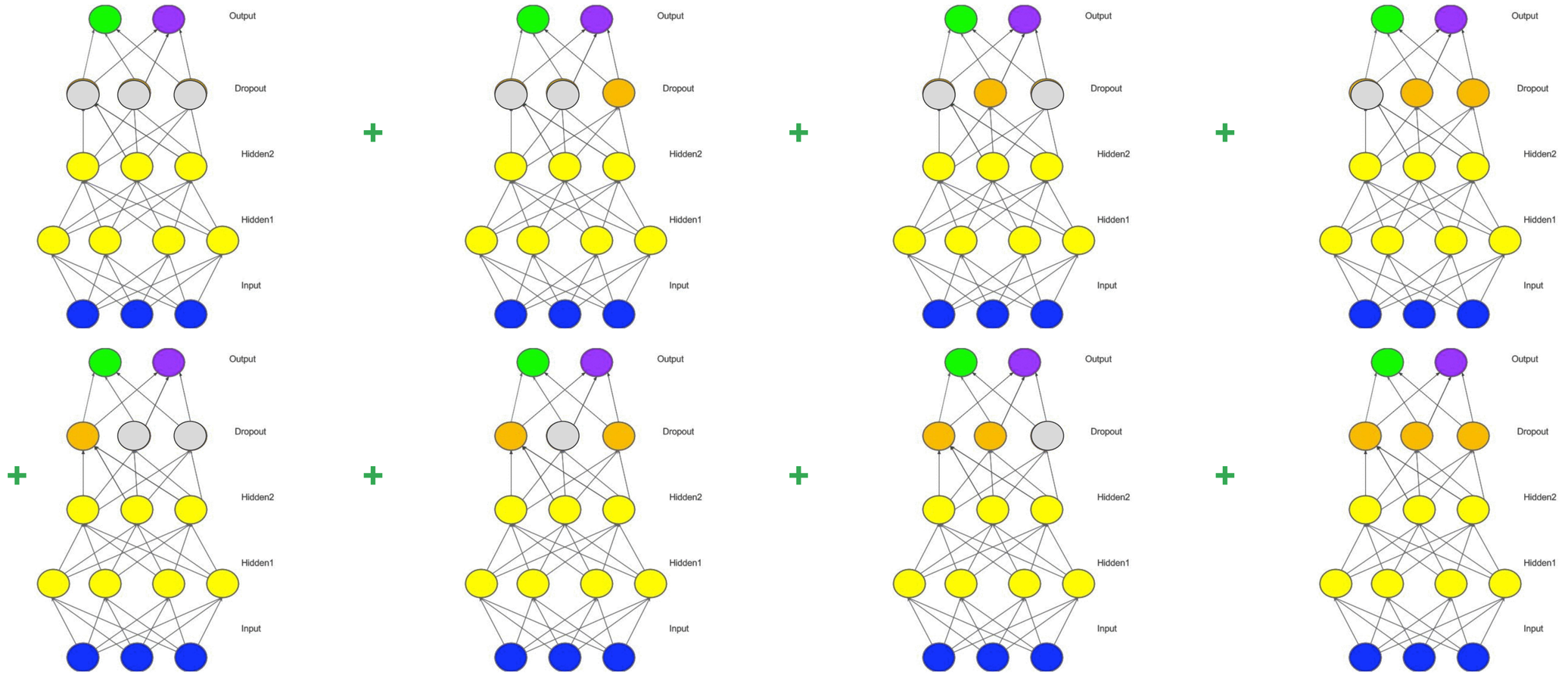
Dropout works by randomly “dropping out” unit activations in a network for a single gradient step.

During training only! In prediction all nodes are kept.

Helps learn “multiple paths” -- think: ensemble models, random forests.



Dropout simulates ensemble learning



The more you drop out, the stronger the regularization

0.0 = no dropout regularization

Intermediate values more useful, a value of dropout=0.2 is typical

The more you drop out, the stronger the regularization

0.0 = no dropout regularization

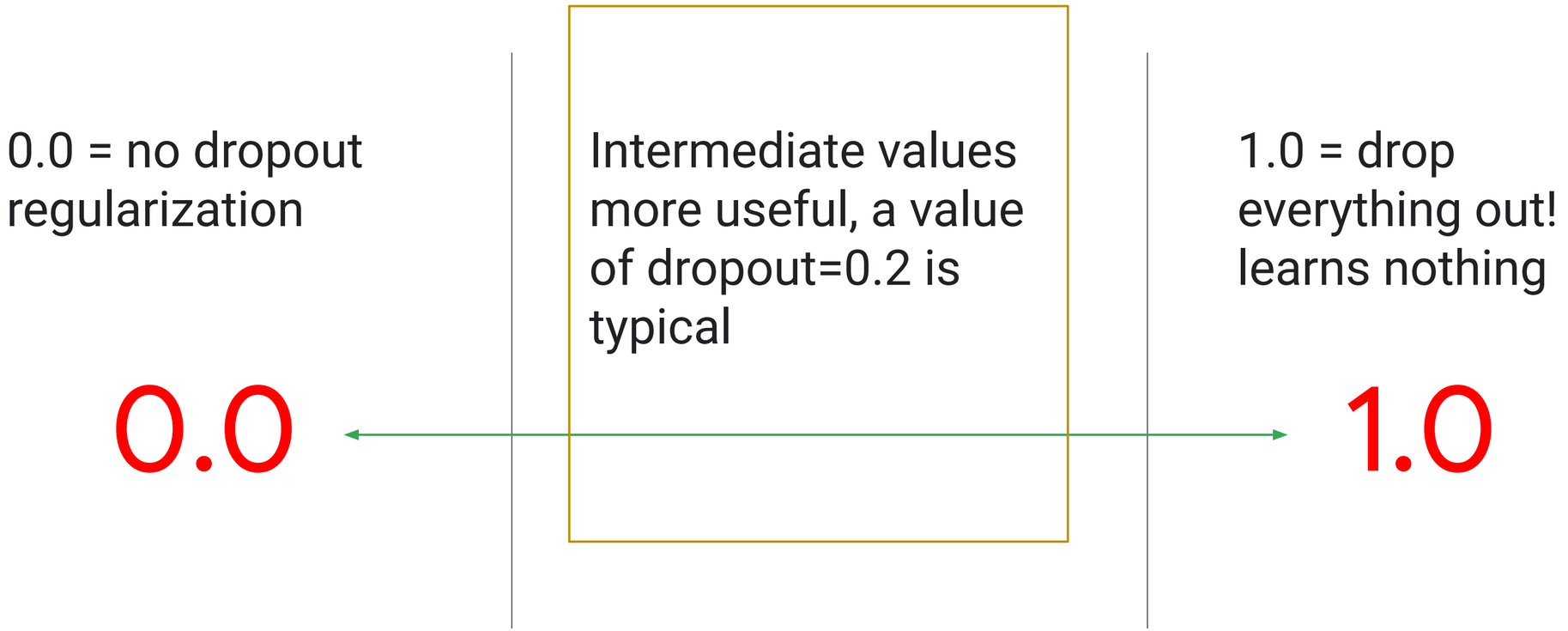
0.0

Intermediate values more useful, a value of dropout=0.2 is typical

1.0 = drop everything out! learns nothing

1.0

The more you drop out, the stronger the regularization



Lab

Keras Functional API

In this lab, you will build a Wide and Deep model to predict the fare amount using Keras functional API

../deepdive2//introduction_to_tensorflow/labs/4_keras_functional_api.ipynb