

AI was used lots throughout this project. Almost only for minor things e.g., correcting an error in the code, inspiration on how to code something suggestions to improve performance and so on. As I think AI's don't perform well with big tasks and most of the time it's easier to do it yourself or find a better solution online.

LLMs used: MistralAI, Gemini2.5 flash, Gemini 2.5 pro, Qwen3, a little bit of copilot.

Bigger tasks AI did:

Gemini2.5 Pro:

Input: "Write me a python function, which is a custom learning rate scheduler for a neural network. It should consist of a linear warmup and a cosine decay."

Output:

```
# Define a custom learning rate schedule class with warmup and cosine decay
class WarmupCosineDecay(tf.keras.optimizers.schedules.LearningRateSchedule):
    """
    A custom learning rate schedule that implements a linear warmup
    followed by a cosine decay.
    """
    def __init__(self, peak_lr, warmup_steps, decay_steps, alpha=0.0, name=None):
        super().__init__()
        self.peak_lr = peak_lr
        self.warmup_steps = warmup_steps
        self.decay_steps = decay_steps
        self.alpha = alpha
        self.name = name

    def __call__(self, step):
        with tf.name_scope(self.name or "WarmupCosineDecay"):
            # Ensure step is a float for calculations
            step = tf.cast(step, tf.float32)

            # --- 1. Warmup Phase ---
            # Linearly increase the learning rate from 0 to peak_lr
            warmup_lr = self.peak_lr * (step / self.warmup_steps)

            # --- 2. Cosine Decay Phase ---
            # Define the cosine decay schedule
            cosine_decay_schedule = tf.keras.optimizers.schedules.CosineDecay(
```

```

        initial\_learning\_rate=self.peak\_lr,
        decay\_steps=self.decay\_steps,
        alpha=self.alpha
    )
    # Calculate the learning rate for the decay phase.
    # Note: The 'step' for the cosine part must be relative to its start.
    decay\_lr = cosine\_decay\_schedule(step - self.warmup\_steps)

    # --- 3. Choose the correct phase ---
    # Use tf.where to select the learning rate based on the current step
    learning\_rate = tf.where(
        step < self.warmup\_steps,
        warmup\_lr,
        decay\_lr
    )
    return learning\_rate

```

google jules:

Input: make a gitignore file for this repository. do not remove the directories "not_relevant", they are relevant. Remove stuff like ipynbcheckpoints or pycache

Output:

```

# Ignore Jupyter Notebook checkpoints
.ipynb\_checkpoints/

```

```

# Ignore Python bytecode
\_\_pycache\_\_/_

```

Input: all the irrelevant files except for .tex, .bib, .pdf, .png, .jpg should be ignored too. also ignore gitattributes

Output:

I've received the latest code review feedback, and I see a potential contradiction in the

The repository's main purpose seems to be machine learning, and we've identified several

If I follow this instruction literally, all the Python and Notebook source files will be

Could you please clarify? Should I:

Proceed with ignoring all files except .tex, .bib, .pdf, .png, and .jpg?

Also include .py and .ipynb files in the whitelist of files to keep?

Take a different approach to the .gitignore file?

Once I have this clarification, I can finalize the .gitignore file correctly.

Input: I meant all the irrelevant files except .tex, .bib, .pdf, .png, and .jpg in the documentation directory.

Output: gitignore files were published in a branch "add-gitignore-rules"

google jules:

Input:update the README.md file

Output:

Update the .gitignore file.

Add entries to ignore .ipynb_checkpoints/, _pycache_/, and .gitattributes.

Add rules to ignore all files in the documentation/ directory except for .tex, .bib, .pdf.

Update the README.md file.

Restructure the README.md to be more organized and detailed.

Add a comprehensive table of contents.

Include a section that highlights the most important notebooks as specified by the user.

Complete pre-commit steps.

Complete pre-commit steps to make sure proper testing, verifications, reviews and reflections.

Submit the changes.

Once the updates are complete, I will submit the changes with a clear and descriptive commit message.

Input:why update the gitignore?

Output:

I'm updating the .gitignore file to incorporate the preferences you shared earlier. Specifically:

Ignore the .ipynb_checkpoints/, _pycache_/, and .gitattributes files.

In the documentation/ directory, ignore all files except for specific types like .tex, .bib, .pdf.

This ensures that the repository remains clean and only tracks the files relevant to the project.

Does the plan I've laid out seem right to you? Once you approve, I'll start working on the updates.

Input:but this already is in the repository.

Output:

You're right, my apologies. I see that .ipynb_checkpoints/ and _pycache_/ are already in the .gitignore file.

My plan was to also add rules to ignore the .gitattributes file, and to set up specific rules for the documentation/ directory.

The full plan remains:

Update the .gitignore file: Add the remaining rules for .gitattributes and the documentation/ directory.

Update the README.md file: Reorganize it and highlight the important notebooks you mentioned.

Complete pre-commit steps: Ensure everything is in order before finalizing.

Submit the changes.

Does this updated plan look good to you?

Input:do not update the gitignore file. Only update the readme.md file

Output:

Update the README.md file.

Restructure the README.md to be more organized and detailed.

Add a comprehensive table of contents.

Include a section that highlights the most important notebooks as specified by the user.

Complete pre-commit steps.

Complete pre-commit steps to make sure proper testing, verifications, reviews and reflections.

Submit the changes.

Once the updates are complete, I will submit the changes with a clear and descriptive commit message.

The README.md file has been updated to be more detailed and structured, including a table of contents.

Input:where did you get this shit with sin(x) function from?? that's not what this repo is for. you are hallucinating.

Output:

I sincerely apologize for the mistake. It seems I was working with incorrect information.

Could you please tell me the correct purpose of this repository? I will update the README.md accordingly.

Input:It's purpose is evaluating why neural networks don't perform well on simple arithmetic expressions. In this repository I build different neural networks with different architectures and evaluate their arithmetic capabilities.

Output:

Thank you for the clarification. I understand now. The project's purpose is to evaluate the performance of different neural network architectures on simple arithmetic expressions.

I have corrected the README.md file to accurately reflect the project's purpose, which is to evaluate the performance of different neural network architectures on simple arithmetic expressions.

Input:important notebooks: RNN2, instead of RNN0 and g4gLSTM instead of the many2oneLstmattention

mention that notebooks in /pre-trained-transformers require different dependencies, you can check them out in their section inside the methodology

also include: FNN1.1 and GetXY script

other important notebooks are: FNN1, RNN0, transformer0

can you make the notebooks and scripts and directories clickable?

Output:

Thank you for the detailed feedback. I will update the README.md with all the changes you requested.

I have revised the README.md file with the specific updates requested by the user, including the addition of a section on dependencies for pre-trained transformers.

I have completed the pre-commit steps, including a code review and recording the latest commit hash.

Input:I like how you did the whole Important files gimmick. Just a small change:
add the transformer4 and transformer 5 to the important files
inside the important files: also add a section with pre-trained fine-tuned trans-
formers pick out 2 important files from there and list them

Output:

Of course. I can make those final adjustments to the README.md file. I'll add the Transf
I have explored the pre-trained-transformers/ directory and selected two important noteb
I have revised the README.md with the final additions, including the new Transformer and
I have completed the final pre-commit steps, including a code review and recording the l
---->>docs-final-readme-update branch

Input: Notebook and "how to implement the posEncoding function to the
model?"

To implement the posEncoding function into your Keras model, you'll need to add it as
Here's how you can modify your build_model function to include the positional encoding.
First, you need to ensure your posEncoding and get_angles functions are correctly defin

```
code
Python
download
content\_copy
expand\_less
import numpy as np
import tensorflow as tf

def get\_angles(pos, i, d\_model):
    angle\_rates = 1 / np.power(10000, (2 * (i // 2)) / np.float32(d\_model))
    return pos * angle\_rates

def posEncoding(max\_seq\_len, d\_model):
    angle\_rads = get\_angles(np.arange(max\_seq\_len)[: , np.newaxis],
                             np.arange(d\_model)[np.newaxis, :],
                             d\_model)

    # Apply sine to even indices in the array; 2i
    angle\_rads[:, 0::2] = np.sin(angle\_rads[:, 0::2])

    # Apply cosine to odd indices in the array; 2i+1
    angle\_rads[:, 1::2] = np.cos(angle\_rads[:, 1::2])

    pos\_encoding = angle\_rads[np.newaxis, ...]

    return tf.cast(pos\_encoding, dtype=tf.float32)
```

Now, you can integrate this into your `build_model` function. A common way to do this is

Here's an updated `build_model` function:

```
code
Python
download
content\_copy
expand\_less
from tensorflow.keras.layers import Input, Flatten, Dense, PReLU, Dropout, Add
from tensorflow.keras.models import Model
from tensorflow.keras import backend as K

def build\_model(hp, input\_shape):
    K.clear\_session()

    # Hyperparameters
    num\_neurons = hp.Int("num\_neurons", 32, 512, step=32)
    num\_layers = hp.Int("num\_layers", 1, 10)
    dropoutTF = hp.Boolean("dropoutT/F")
    dropout\_rate = hp.Float("dropout\_rate", 0.1, 0.5, step=0.1)

    # Input Layer
    inputs = Input(shape=input\_shape)

    # Assuming input\_shape is (sequence\_length, feature\_dimension)
    # If your input is already flat, you'll need to reshape it to add positional encoding
    sequence\_length = input\_shape[0]
    d\_model = input\_shape[1] if len(input\_shape) > 1 else 1

    # Add positional encoding
    positional\_encoding = posEncoding(sequence\_length, d\_model)
    x = Add()([inputs, positional\_encoding])

    x = Flatten()(x)

    # Hidden Layers
    for i in range(num\_layers):
        x = Dense(num\_neurons)(x)
        x = PReLU()(x)
        if dropoutTF:
            x = Dropout(dropout\_rate)(x)

    # Output Layer
    outputs = Dense(1, activation='linear')(x)
```

```
# Create and compile the model
lil\_model = Model(inputs=inputs, outputs=outputs)
lil\_model.compile(optimizer="adam", loss="mse")

return lil\_model
```

Explanation of Changes and How to Use It:

Functional API: I've switched to the Keras Functional API (Input and Model) instead of Sequential.

Determining sequence_length and d_model: The posEncoding function requires the maximum sequence length and the embedding dimension.

Adding Positional Encoding:

positional_encoding = posEncoding(sequence_length, d_model): This creates the positional encoding matrix.

x = Add()([inputs, positional_encoding]): The positional encodings are added directly to the input embeddings.

Flatten Layer: After adding the positional encoding, the data is flattened before being passed to the next layer.

To use this new build_model function, you just need to pass the input_shape as you were previously.

Important Considerations:

Input Shape: The most critical part is ensuring input_shape is correctly interpreted. Pay attention to the order of dimensions.

Broadcasting: The Add layer will broadcast the positional encoding matrix to match the shape of the input embeddings.

Input: "Input 0 of layer "dense" is incompatible with the layer: expected axis -1 of input shape to have value 225, but received input with shape (None, 15)
 Arguments received by Functional.call(): • inputs=tf.Tensor(shape=(None, 15), dtype=float32) • training=True • mask=None • kwargs={}

You've run into a very common and insightful error in Keras. Let's break down what's happening.

Understanding the Error

The error message "Input 0 of layer "dense" is incompatible with the layer: expected axis -1 of input shape to have value 225, but received input with shape (None, 15)" indicates a mismatch between the expected and actual input shapes.

What the model received: During training or prediction (.fit() or .predict()), the first Dense layer received an input with a shape of (None, 15).

What the model expected: When the model was built, the first Dense layer was somehow configured to expect an input with a shape of (None, 225).

The root cause of this mismatch is a shape incompatibility when trying to add the positional encoding to the input embeddings.

The Solution: Reshaping the Input

To make this work, we need to treat your input vector of 15 features as a sequence of 15

Here is the corrected build_model function that properly integrates the positional encoding

```
code
Python
download
content\_copy
expand\_less
import numpy as np
import tensorflow as tf
from tensorflow.keras.layers import Input, Flatten, Dense, PReLU, Dropout, Add, Reshape
from tensorflow.keras.models import Model
from tensorflow.keras import backend as K

# --- Make sure these functions are defined in your notebook ---
def get\_angles(pos, i, d\_model):
    angle\_rates = 1 / np.power(10000, (2 * (i // 2)) / np.float32(d\_model))
    return pos * angle\_rates

def posEncoding(max\_seq\_len, d\_model):
    angle\_rads = get\_angles(np.arange(max\_seq\_len)[: , np.newaxis],
                             np.arange(d\_model)[np.newaxis, :],
                             d\_model)
    # Apply sine to even indices
    angle\_rads[:, 0::2] = np.sin(angle\_rads[:, 0::2])
    # Apply cosine to odd indices
    angle\_rads[:, 1::2] = np.cos(angle\_rads[:, 1::2])
    pos\_encoding = angle\_rads[np.newaxis, ...]
    return tf.cast(pos\_encoding, dtype=tf.float32)

# --- Updated build\_model function ---
def build\_model(hp, input\_shape):
    K.clear\_session()

    # Hyperparameters
    num\_neurons = hp.Int("num\_neurons", 1, 512)
    num\_layers = hp.Int("num\_layers", 1, 16)
    dropoutTF = hp.Boolean("dropoutT/F")

    # --- Model Architecture ---
    inputs = Input(shape=input\_shape) # Expects shape like (15,)

    # The input shape is (batch, 15). We need to treat it as a sequence.
    # Reshape to (batch, 15, 1) to represent a sequence of 15 steps with 1 feature each.
```

```

# This is the key step to make shapes compatible.
sequence\_length = input\_shape[0]
d\_model = 1
x = Reshape((sequence\_length, d\_model))(inputs)

# Add positional encoding
positional\_encoding = posEncoding(sequence\_length, d\_model)
x = Add()(x, positional\_encoding])

# Flatten the sequence back into a vector of shape (batch, 15) for the Dense layers.
x = Flatten()(x)

# Add the dense layers as before
for i in range(num\_layers):
    x = Dense(num\_neurons)(x)
    x = PReLU()(x)
    if dropoutTF:
        x = Dropout(0.1)(x)

# Output Layer
outputs = Dense(1, activation='linear')(x)

# Create and compile the model
model = Model(inputs=inputs, outputs=outputs)
model.compile(optimizer="adam", loss="mse")

return model

```

You can now use this build_model function with your Keras Tuner
 # Re-run the tuner search after updating the function.
 How to Use the New Function

Replace the old build_model function in your notebook with the new, corrected version a

Make sure the get_angles and posEncoding functions are also present in a cell before bu

Re-run the Keras Tuner search. Because you've changed the model's architecture by adding

```

code
Python
download
content\_copy
expand\_less
# Re-initialize the tuner and run the search again
tuner = keras\_tuner.BayesianOptimization(
    hypermodel=lambda hp: build\_model(hp, input\_shape),

```

```

    # ... other tuner parameters ...
)

tuner.search(train\_dataset, epochs=num\_epochs, validation\_data=val\_dataset)

```

Input: "there is a lot of code, that isn't necessary in this fle. make this file shorter." (attached file)

Of course, the provided Python script contains a significant amount of code for training

Key Changes:

Removed Data Loading: All code related to loading and batching `x_train`, `y_train`, `x_val`

Removed Model Training and Tuning: The `keras_tuner` implementation, including the `build_model`

Removed Evaluation Metrics: All sections calculating Mean Absolute Error (MAE) and Mean

Combined Model Creation and Loading: A new function, `create_and_load_model`, is introduced

Kept Essential Components: The script retains the necessary imports, the path additions

Shortened `FNN5reciever.py`

code

Python

download

`content_copy`

`expand_less`

`import sys`

`import os`

`import tensorflow as tf`

`from tensorflow.keras.layers import Dense, PReLU, Dropout, Flatten`

`from tensorflow.keras.models import Sequential`

`from tensorflow.keras import Input`

`# --- Setup Project Path ---`

`# This allows the script to find the custom 'GetXY' module in the parent directory.`

`try:`

`# Get the absolute path of the current script's directory`

`current_dir = os.path.dirname(os.path.abspath(__file__))`

`except NameError:`

`# Fallback for interactive environments like Jupyter`

`current_dir = os.path.dirname(os.path.abspath("FNN5reciever.py"))`

`parent_dir = os.path.dirname(current_dir)`

`sys.path.append(parent_dir)`

```

# --- Import Custom Tokenizer ---
# This tokenizer is required to preprocess user input.
from GetXY import tokenizer

def create\_and\_load\_model(weights\_path):
    """
    Defines the model architecture, compiles it, and loads the pre-trained weights.

    NOTE: You must define the model with the *exact* same architecture and
    hyperparameters as the model that was saved. Adjust the values below if needed.
    """
    # Define the model's input shape. This should match the shape of the tokenized input
    # Example: if your tokenizer creates vectors of length 50, shape would be (50,)
    input\_shape = (15,) # <--- IMPORTANT: SET THIS TO MATCH YOUR TOKENIZER'S OUTPUT SHAPE

    # Best hyperparameters found during tuning (replace with your actual best values)
    best\_hps = {
        "num\_layers": 8,
        "num\_neurons": 269,
        "dropout": True,
        "dropout\_rate": 0.1
    }

    # Re-create the model with the best hyperparameters
    model = Sequential()
    model.add(Input(shape=input\_shape))
    model.add(Flatten())

    for \_ in range(best\_hps["num\_layers"]):
        model.add(Dense(best\_hps["num\_neurons"]))
        model.add(PReLU())
        if best\_hps["dropout"]:
            model.add(Dropout(best\_hps["dropout\_rate"]))

    model.add(Dense(1, activation='linear'))

    # Compile the model (optimizer and loss are needed for loading, but not for inference)
    model.compile(optimizer="adam", loss="mse")

    # Load the trained weights
    model.load\_weights(weights\_path)

    print("Model architecture created and weights loaded successfully.")
    model.summary()

```

```

        return model

# --- Main Application ---
if \_name\_ == "\_main\_":
    # Path to the saved model weights
    WEIGHTS\_FILE = "FNN2\_weights.weights.h5"

    # Create the model and load the weights
    best\_model = create\_and\_load\_model(WEIGHTS\_FILE)

    # --- Interactive Calculator Loop ---
    print("\n-----")
    print("HELLO AND WELCOME TO THE NEURAL PREDICTIVE CALCULATOR - NPC")
    print("-----")

    while True:
        try:
            user\_expr = input("What do you want me to calculate? (or type 'exit' to quit): ")
            if user\_expr.lower() == 'exit':
                break

            # The tokenizer converts the string expression into a numerical vector
            tokenized\_input = tokenizer([user\_expr])

            # Use the model to predict the result
            model\_pred = best\_model.predict(tokenized\_input)

            print(f"what you asked: {user\_expr}")
            print(f"This should be the answer: {model\_pred[0][0]:.4f}")
            print("-----")

        except Exception as e:
            print(f"An error occurred: {e}")
            print("Please ensure your input is a valid mathematical expression.")
            print("-----")

```

Input: create one more cell in the transformer5 notebook, where it uses the same logic as in FNN6.py to calculate the pvalue among others. answer with the code.

Output:

```

#this was generated based on the FNN6.py script, which was made by myself. (Gemini 2

import numpy as np
from scipy.stats import ttest\_1samp

# Initialize lists to store metrics for each bootstrap run

```

```

MAEinRange\_list = []
MREinRange\_list = []
MAEoutRange\_list = []
MREoutRange\_list = []
MAElongRange\_list = []
benchmarks\_list = []

# Loop through the predictions from each bootstrap run
for i in range(n\_bootstrap):
    # Extract predictions for the current run
    current\_preds\_in\_range = bootstrap\_predsInRange[i]
    current\_preds\_out\_range = bootstrap\_predsOutOfRange[i]
    current\_preds\_long\_range = bootstrap\_predsLongRange[i]

    # --- In-Range Metrics ---
    diffInRange = np.abs(y\_test - current\_preds\_in\_range.flatten())
    safe\_y\_test = np.where(np.isclose(y\_test, 0.0), 1.0, y\_test)
    reldiffInRange = diffInRange / np.abs(safe\_y\_test)

    mean\_mae\_in\_range = np.mean(diffInRange)
    mean\_mre\_in\_range = np.mean(reldiffInRange)
    MAEinRange\_list.append(mean\_mae\_in\_range)
    MREinRange\_list.append(mean\_mre\_in\_range)

    # --- Out-of-Range Metrics ---
    diffOutOfRange = np.abs(out\_y\_test - current\_preds\_out\_range.flatten())
    safe\_out\_y\_test = np.where(np.isclose(out\_y\_test, 0.0), 1.0, out\_y\_test)
    reldiffOutOfRange = diffOutOfRange / np.abs(safe\_out\_y\_test)

    mean\_mae\_out\_range = np.mean(diffOutOfRange)
    mean\_mre\_out\_range = np.mean(reldiffOutOfRange)
    MAEoutRange\_list.append(mean\_mae\_out\_range)
    MREoutRange\_list.append(mean\_mre\_out\_range)

    # --- Long-Range Metrics ---
    diffLongRange = np.abs(long\_y\_test - current\_preds\_long\_range.flatten())
    mean\_mae\_long\_range = np.mean(diffLongRange)
    MAElongRange\_list.append(mean\_mae\_long\_range)

    # --- Benchmark Calculation ---
    # This part replicates the specific slicing and filtering from FNN6.py for the benchmark

    # Specific filter for out-of-range benchmark
    placeholder = absSum(outsideExpr)
    indices\_with\_placeholder\_22 = [i for i, val in enumerate(placeholder) if val == 2]

```

```

diffOutOfRange\_for\_benchmark = []
for idx in indices\_with\_placeholder\_22:
    diffOutOfRange\_for\_benchmark.append(np.abs(out\_y\_test[idx] - current\_preds\_out\_y))

meanDiff\_OutOfRange\_for\_benchmark = np.mean(diffOutOfRange\_for\_benchmark)

# Specific slice for long-range benchmark
diffLongRange\_for\_benchmark = []
for j in range(200, 300):
    diffLongRange\_for\_benchmark.append(np.abs(long\_y\_test[j] - current\_preds\_long\_y))

meanDiff\_LongRange\_for\_benchmark = np.mean(diffLongRange\_for\_benchmark)

# Calculate the benchmark score for the current run
benchmark = 0
benchmark += baseline\_deviation / (mean\_mae\_in\_range**2) / 4
benchmark += baseline\_out\_deviation / (meanDiff\_OutOfRange\_for\_benchmark**2) / 4
benchmark += baseline\_long\_deviation / (meanDiff\_LongRange\_for\_benchmark**2) / 4
benchmark += baseline\_relError / (mean\_mre\_out\_range**2) / 4
benchmarks\_list.append(benchmark)

# --- Statistical Analysis and Final Output ---

# Perform one-sample t-test against a population mean of 1
# Note: A t-test is meaningful only if n\_bootstrap > 1
if n\_bootstrap > 1:
    stats1, p\_value1 = ttest\_1samp(MAEinRange\_list, popmean=1)
    stats2, p\_value2 = ttest\_1samp(MREinRange\_list, popmean=1)
    stats3, p\_value3 = ttest\_1samp(MAEoutRange\_list, popmean=1)
    stats4, p\_value4 = ttest\_1samp(MREoutRange\_list, popmean=1)
    stats5, p\_value5 = ttest\_1samp(MAElongRange\_list, popmean=1)
    stats6, p\_value6 = ttest\_1samp(benchmarks\_list, popmean=1)

    print(f"MAE in Range P-value: {p\_value1}")
    print(f"MRE in Range P-value: {p\_value2}")
    print(f"MAE out Range P-value: {p\_value3}")
    print(f"MRE out Range P-value: {p\_value4}")
    print(f"MAE long Range P-value: {p\_value5}")
    print(f"benchmark P-value: {p\_value6}\n")
else:
    print("Cannot calculate p-values with n\_bootstrap=1. Run more bootstraps for statistics")

# Print average metrics across all runs
print(f"Average MAE in Range: {np.mean(MAEinRange\_list)}")
print(f"Average MRE in Range: {np.mean(MREinRange\_list)}")

```

```

print(f"Average MAE out Range: {np.mean(MAEoutRange\_list)}")
print(f"Average MRE out Range: {np.mean(MREoutRange\_list)}")
print(f"Average MAE long Range: {np.mean(MAElongRange\_list)}")
print(f"Average benchmark: {np.mean(benchmarks\_list)}\n")

```

```

# Print the lists of metrics
print(f"MAE in Range List: {MAEinRange\_list}")
print(f"MRE in Range List: {MREinRange\_list}")
print(f"MAE out Range List: {MAEoutRange\_list}")
print(f"MRE out Range List: {MREoutRange\_list}")
print(f"MAE long Range List: {MAElongRange\_list}")
print(f"Benchmark List: {benchmarks\_list}")

```

Input: loss landscape of a neural network, with the number of neurons and number of layers being the two axis of the plain. The plain caves in at multiple sweetspots. The height is loss. draw it in matplotlib.pyplot. add some randomness to the mix, so the landscape isn't smooth. The landscape goes from 1 to 512 neurons and from 1 to 16 layers. here are the values of the sweetspots: (only use num_neurons and num_layers, ignore the rest.)

Trial 025 summary Hyperparameters: num_neurons: 345 num_layers: 1 dropoutT/F: False Score: 0.0012616724707186222

Trial 000 summary Hyperparameters: num_neurons: 348 num_layers: 1 dropoutT/F: False Score: 0.001294966321438551

Trial 056 summary Hyperparameters: num_neurons: 439 num_layers: 4 dropoutT/F: False Score: 0.0025118887424468994

Trial 020 summary Hyperparameters: num_neurons: 16 num_layers: 1 dropoutT/F: False Score: 0.002628110349178314

Trial 026 summary Hyperparameters: num_neurons: 289 num_layers: 2 dropoutT/F: False Score: 0.002691656118258834

Trial 042 summary Hyperparameters: num_neurons: 435 num_layers: 4 dropoutT/F: False Score: 0.0027040746062994003

Trial 094 summary Hyperparameters: num_neurons: 440 num_layers: 4 dropoutT/F: False Score: 0.002829659730195999

Trial 064 summary Hyperparameters: num_neurons: 440 num_layers: 4 dropoutT/F: False Score: 0.0029272951651364565

Trial 044 summary Hyperparameters: num_neurons: 435 num_layers: 4 dropoutT/F: False Score: 0.003144865622743964

Trial 008 summary Hyperparameters: num_neurons: 200 num_layers: 2 dropoutT/F: False Score: 0.0031737370882183313

Output:

I'll create a 3D loss landscape visualization with valleys at your specified sweetspots

```

'''python
import numpy as np
import matplotlib.pyplot as plt

```

```

from mpl\toolkits.mplot3d import Axes3D
from scipy.ndimage import gaussian\_filter

# Define the grid
neurons = np.linspace(1, 512, 100)
layers = np.linspace(1, 16, 50)
X, Y = np.meshgrid(neurons, layers)

# Define sweetspots from your data
sweetspots = [
    (345, 1, 0.00126),
    (348, 1, 0.00129),
    (439, 4, 0.00251),
    (16, 1, 0.00263),
    (289, 2, 0.00269),
    (435, 4, 0.00270),
    (440, 4, 0.00283),
    (440, 4, 0.00293), # duplicate location
    (435, 4, 0.00314), # duplicate location
    (200, 2, 0.00317),
]

# Create base landscape with slight upward trend for complexity
# Higher loss for very small and very large networks
Z = 0.01 + 0.002 * np.log1p(X/100) + 0.001 * np.log1p(Y/4)
Z += 0.003 * np.exp(-X/200) # Penalty for very small networks
Z += 0.002 * np.exp((X-512)/200) # Penalty for very large networks

# Add random noise for rough texture
np.random.seed(42)
noise = np.random.randn(*Z.shape) * 0.0008
Z += noise

# Create depressions at sweetspots
for neurons\_val, layers\_val, loss\_val in sweetspots:
    # Create Gaussian depression centered at each sweetspot
    sigma\_n = 40 # width in neuron dimension
    sigma\_l = 1.5 # width in layer dimension

    # Calculate distance from this sweetspot
    dist\_n = (X - neurons\_val)**2 / (2 * sigma\_n**2)
    dist\_l = (Y - layers\_val)**2 / (2 * sigma\_l**2)

    # Create depression (lower values at sweetspot)
    depression = np.exp(-(dist\_n + dist\_l))

```

```

    # Blend the depression with target loss value
    Z = Z * (1 - 0.8 * depression) + loss\_val * depression

# Apply mild smoothing to blend features while keeping roughness
Z = gaussian\_filter(Z, sigma=0.5)

# Add final layer of small random noise
Z += np.random.randn(*Z.shape) * 0.0002

# Create the 3D plot
fig = plt.figure(figsize=(14, 10))
ax = fig.add\_subplot(111, projection='3d')

# Plot the surface
surf = ax.plot\_surface(X, Y, Z, cmap='viridis',
                        edgecolor='none', alpha=0.9,
                        antialiased=True, shade=True)

# Mark the sweetspots with red dots
for neurons\_val, layers\_val, loss\_val in sweetspots:
    ax.scatter([neurons\_val], [layers\_val], [loss\_val],
               color='red', s=50, alpha=0.8,
               edgecolors='darkred', linewidth=1)

# Labels and title
ax.set\_xlabel('Number of Neurons', fontsize=12, labelpad=10)
ax.set\_ylabel('Number of Layers', fontsize=12, labelpad=10)
ax.set\_zlabel('Loss', fontsize=12, labelpad=10)
ax.set\_title('Neural Network Loss Landscape\n(Red dots indicate optimal configurations)
             fontsize=14, pad=20)

# Set viewing angle for better visualization
ax.view\_init(elev=25, azim=45)

# Add colorbar
fig.colorbar(surf, ax=ax, shrink=0.5, aspect=5, pad=0.1)

# Set axis limits
ax.set\_xlim(1, 512)
ax.set\_ylim(1, 16)

plt.tight\_layout()
plt.show()

# Create a contour plot view from above
fig2, ax2 = plt.subplots(figsize=(12, 8))

```

```

contour = ax2.contourf(X, Y, Z, levels=20, cmap='viridis', alpha=0.8)
contour\_lines = ax2.contour(X, Y, Z, levels=15, colors='black', alpha=0.3, linewidths=0)

# Mark sweetspots
for neurons\_val, layers\_val, loss\_val in sweetspots:
    ax2.scatter(neurons\_val, layers\_val, color='red', s=100,
                edgecolors='darkred', linewidth=2, zorder=5)
    ax2.annotate(f'{loss\_val:.4f}', (neurons\_val, layers\_val),
                xytext=(5, 5), textcoords='offset points',
                fontsize=8, color='darkred')

ax2.set\_xlabel('Number of Neurons', fontsize=12)
ax2.set\_ylabel('Number of Layers', fontsize=12)
ax2.set\_title('Loss Landscape - Top View\n(Red dots with values show optimal configurations)',
              fontsize=14)
ax2.grid(True, alpha=0.3)

plt.colorbar(contour, ax=ax2, label='Loss')
plt.tight\_layout()
plt.show()
'''

```

This code creates two visualizations:

1. ****3D Surface Plot****: Shows the loss landscape as a 3D surface with:
 - Valleys (depressions) at your specified sweetspot locations
 - Random roughness throughout the surface
 - Red dots marking the exact optimal configurations
 - The surface shows how loss varies with both the number of neurons and layers
2. ****Contour Plot****: A top-down view showing:
 - The same landscape as a heat map
 - Contour lines showing loss levels
 - Red dots with annotations showing the exact loss values at sweetspots

The landscape features:

- Lower loss values (deeper valleys) at your specified optimal configurations
- General trend showing that very small or very large networks tend to have higher loss
- Random variations to create a realistic, non-smooth surface
- Clear visualization of multiple local minima in the hyperparameter space

The sweetspots around (345-348 neurons, 1 layer) and (435-440 neurons, 4 layers) appear

1 AI assistance with grammar errors on documentation documents.

Input: "Correct the attached document for grammar errors and give me feedback. Latex code for document:"

Output: *Corrections for the document*