**Problem 1**

1. **Increasing the feature weight of 'circle'**: When you increase the weight of a particular feature (like 'circle'), it will likely increase the predicted counts for objects that include that feature. In this case, if 'circle' is a significant feature for certain objects, increasing its weight will likely increase the expected counts for those objects that have a circle in them.
2. **Show Gradient**: The red and blue bars under feature weights might represent the positive and negative gradients respectively. Red could indicate an increase in the expected counts with an increase in the feature weight, while blue might signify a decrease.
3. **Values of features when 'Solve' is clicked**: Unfortunately, I can't directly view or interact with external content like screenshots. But typically, when you click 'Solve', it should display the optimized values of the feature weights that best fit the model to the observed data.
4. **Regularization effects**: When regularization is applied (like L1 regularization), it penalizes large weights, potentially forcing some weights towards zero. This is why the values of features might differ between regularized and non-regularized models.
5. **Dividing by Z and expression of Z**: Dividing by Z (the partition function) is necessary to ensure that the probabilities of all possible outcomes sum up to 1, making it a valid probability distribution. The expression for Z in the context of a softmax function ensures normalization across all possible outcomes:

Z =∑ all outcomes \* e ^(feature weights)

This way, dividing each term by Z ensures that the sum of probabilities of all outcomes equals 1, conforming to the requirements of a probability distribution.

**Problem 2**

**Best features for separation:**

The features that might give the best separation in this dataset are nonlinear transformations or combinations of the original features. These can help in creating decision boundaries that effectively separate the two spirals. You'd want to observe the interplay of different combinations or transformations of features to find those that result in clearer boundaries between the spirals.

**Regularizer effects:**

When a regularizer is added, such as L1 or L2 regularization, it introduces constraints on the model weights. In the context of the spiral dataset, adding a regularizer might impact the complexity of the model. This could lead to a situation where more epochs are required to converge because the regularization is imposing constraints on the model's learning.

**Separation Visualization:**

You can adjust the features in the Tensorboard Playground to find the combination that offers the best separation. After experimenting with different feature transformations or combinations, the visualization will demonstrate the resulting decision boundaries, showcasing how well the model separates the two spirals.

**Problem 3**

**Observing Layer Shapes for Kernel Size, Stride Length, and Maxpool Window Size:**

* In CNNs, the shapes of each layer in the model architecture often reveal information about the kernel size, stride length, and max-pooling window size. For instance, a shape like (3, 3, 32) indicates a kernel size of 3x3 with 32 filters in that layer.

**Image Upload and Observations:**

1. **Uploading the Rattlesnake Image:**
   * Upload the image of the rattlesnake into the CNN Explainer.
   * Observe the output generated by the model after analyzing the rattlesnake image. Take note of the class prediction or any other relevant information provided by the model.
2. **Uploading the Coyote Image:**
   * Similarly, upload the image of the coyote into the CNN Explainer.
   * Note the output provided by the model after analyzing the coyote image. Compare it to the prediction for the rattlesnake image and look for any differences or similarities in the model's predictions.

**Class 'Espresso':**

* Analyze how the CNN Explainer classifies the images of 'Espresso.' If the model consistently misclassifies or struggles with the class 'Espresso,' it might suggest that the model has difficulty distinguishing or identifying this particular class. This could indicate potential challenges or ambiguities in the dataset or similarity with other classes.