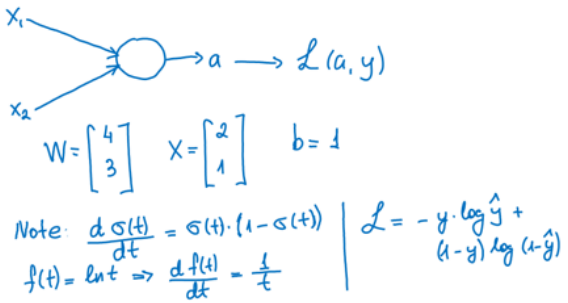
1. Below is a single neuron for cat classification problem (logistic regression) with it's input X, parameters W, b and learning rate 0.5. What will be the updated values of W and b after first forward-backward propagation if an input was misclassified i.e. then input was predicted as 'cat' while it was actually 'non-cat'? (10pt)



2. Explain "No free lunch theorem". (7pt)

3. What is the intuition behind "dropout" technique? (8pt)

4. Explain the following optimization algorithm: RMSprop. (9pt)

5. Consider the following code snippet: (5pt)

```
# a.shape = (3,4)
# b.shape = (4,1)
for i in range(3):
  for j in range(4):
    c[i][j] = a[i][j] + b[j]
```

How would you vectorize it?

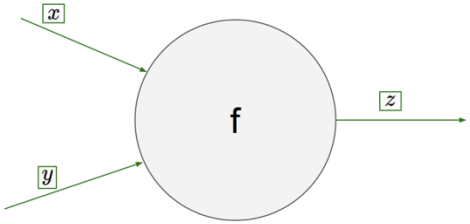
6. Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.) (6pt)
   - Exploding gradients
   - Vanishing gradients
   - Xavier initialization
   - Dropout
   - L2 regularization
   - Data augmentation
   - RMSprop
   - Batch normalization

7. Difference between parameters and hyperparameters. Provide examples. (5pt)

8. MLP's with one (large) hidden layer are universal function approximators already. Why do we want to use deeper architectures? (7pt)

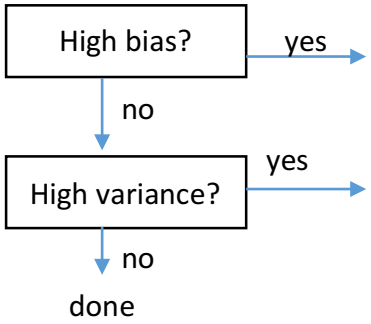
9. Explain 2 problems of sigmoid activation function (8pt)
   - Saturated neurons "kills" the gradietns:
   - Sigmoid outputs are not zero centered:

10. Given forward flow of the variables below, show their corresponding backward flow. (7pt)



11. Why would we need batch normalization? Explain it's benefits. (10pt)

12. Basic recipe for training deep learning models (10pt)



13. Explain cyclical learning rate scheduler. Why it works? How to choose safe range? (8pt)