1. B
2. Accuracy, precision, and recall are good metrics to use in classification for balanced datasets. For imbalanced data, it is better to use ROC/AUC. I would first determine what the distribution of the data is. A confusion matrix is an important visual to show various metrics such as precision, recall, ROC/AUC, and F1 score. Start with accuracy, and then do a deeper dive with the confusion matrix to look at the precision and recall. Then, if we see a trend where precision or recall has an extreme value, F1 score would help balancing both precision and recall. Finally, an ROC would show the relationship between sensitivity and specificity. A high AUC value would be preferable for our model.
3. For gradient descent, there are many appropriate selections for model criteria. In some cases, the stop points for gradient descent to stop running is when step size approaches zero. In other cases, limiting it based off a fixed number of iterations or simply minimizing the MSE. The error can be defined as the MSE or cost function.
4. Attached in github link: https://github.com/DanielKim512/Intro2DL.git
5. a) ROCs with equal AUC but different shapes then depend on certain situations that either value FPR or TPR. In that sense, it is difficult to outright say that one classifier is better than the other without additional context.
b) A low false positive rate is valued in loan situations, which would consider A to be the better ROC curve. This is because a false positive would mean a bad customer would receive a loan that they would not be able to repay.
c) A high sensitivity, true positive rate, is more valued in medical tests, then curve B would be better. This is because showing a false positive on a test can be followed up with additional tests. However, a false negative test would completely miss a diagnosis and lead to a life at stake.
d) The AUC value can be interpreted as the probability that a randomly chosen positive instance will be ranked higher than a randomly chosen negative instance. This is equivalent to the Mann-Whitney U test which requires a measure of the central tendencies of the positive and negative instance, sample statistic, the sample size, and significance level.
6. During Backpropagation computations, the weights are varying while the input and output is fixed. Because of the varying weights, while it is possible to compute using the chain rule, separately calculating is inefficient. Therefore, backpropagation is more efficient since it avoids intermediate values and just calculates the gradient of each layer.
7. a) The red ROC is the best because it is giving a perfect classification model, blue ROC is decent but not perfect, and the green ROC is a useless model representing that one has no better odds than chance alone.
b) The ROCs are plotted based off the (1-Specificity) or the false positive rate (FPR) vs the sensitivity or the true positive rate (TPR). The x axis represents the FPR while the y axis represents the TPR. Therefore, for the red ROC, which is hugging the axis, the FPR is 0 while the TPR is 1. For the blue ROC, the FPR hugs 0 but increases slightly indicating that the model did classify false positives. In addition, the TPR is still high but not 1. Finally, the green ROC is simply a linear line indicating that FPR and TPR is linearly increasing.