Cross-entropy loss measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label.

For testing this exercise LR= 0.001,0.0001,0.1;

Chart

Description automatically generated

Fig a: Loss and accuracy when the model is using different learning rates (SGD).

1. Also, Fig a- for sgd optimizer graph shows that at higher learning rate accuracy is higher and risk is low.
2. When we compare both optimizer for same learning rate we can see that adam is performing really well over sgd.
3. Risk analysis: as LR is increasing (0.0001—0.001—0.1) risk is gradually decreasing. Compared to other two, risk is very minimal when lr = 0.1. When lr=0.001 risk is like a curve and for last epochs it is constant. But when lr=0.0001 risk is pretty high also it is not decreasing much at the end.
4. Accuracy Analysis: as LR is increasing (0.0001—0.001—0.1), accuracy is increasing. Compared to other two lr, accuracy is highest when lr=0.1. For lr=0.0001 accuracy is increasing gradually for first few epochs but later it is almost constant to 82 percent. When lr = 0.001 accuracy is increasing for first epochs and then rate is slow but at the end it made it to 96 percent accuracy.
5. When lr=0.001, optimizer=SGD we get accuracy from 20 percent to 96 percent whereas risk is decreasing gradually(from 1.1 to 0.3).
6. When lr=0.0001, optimizer=SGD we get accuracy from 20 percent to 82 percent whereas risk is decreasing slowly(0.6 at the end).
7. When lr=0.1, optimizer=SGD we get accuracy from 20 percent to 98 percent whereas risk is decreasing gradually (0.05 at the end).

Graphical user interface, chart, line chart

Description automatically generated

Fig b: Loss and accuracy when the model is using different learning rates (ADM).

1. Risk analysis: as LR is increasing (0.0001—0.001—0.1) risk is gradually decreasing. Compared to other two, risk is very minimal when lr = 0.1. When lr=0.001 risk is like a curve and for last epochs it is constant. But when lr=0.0001 risk is pretty high also it is not decreasing much at the end.
2. Accuracy Analysis: as LR is increasing (0.0001—0.001—0.1), accuracy is increasing. Compared to other two lr, accuracy is highest when lr=0.1. For lr=0.0001 accuracy is increasing gradually for first few epochs but later it is slow. When lr = 0.001 accuracy is increasing for first epochs and then rate is slow but at the end it made it to 98 percent accuracy.
3. When lr=0.001, optimizer=ADM we get accuracy from 20 percent to 98 percent whereas risk is decreasing gradually(from 1.1 to 0.05).
4. When lr=0.0001, optimizer=ADM we get accuracy from 20 percent to 94 percent whereas risk is decreasing slowly (from 1.1 to 0.2).
5. When lr=0.1, optimizer=ADM we get accuracy from 20 percent to 98 percent whereas risk is decreasing slowly (0.03 at the end).

From above explanation, by looking at number of accuracy and risk: as learning rate is increasing accuracy is increasing. Also ADM optimizer is performing well, giving more accuracy and less risk than SGD.