[민호] 재구현 1

Baseline Network

10 epoch후, training accuracy = 0.8538과 validation accuracy = 0.7017

Testing accuracy = 0.6996

tester

Summary

Model	Attack Success	Hyperparameters
FGSM	0.1348	e = 0.01
Basic Iterative FGSM	0.1482	e= 0.01, a = 0.1, iteration = 5
Least-Likely Iterative FGSM	0.0228	e=0.01, 0.1, iteration =5
Untitled		

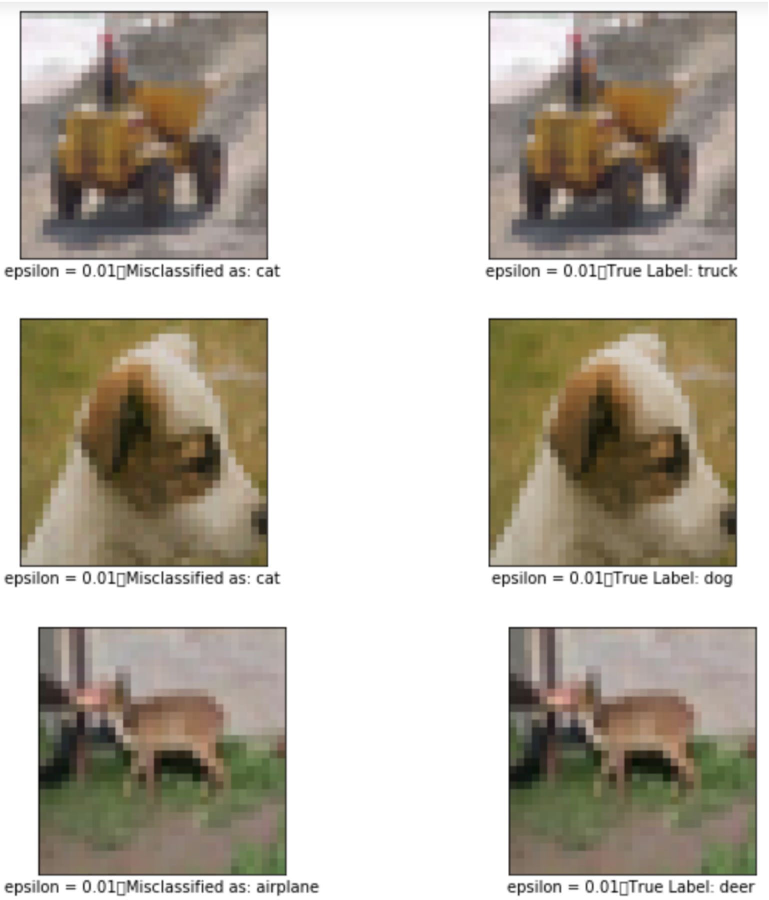
FGSM Method

attack success for different epsilon

Epsilon Value	Accuracy
0.01	0.1348
0.05	0.1629
0.10	0.1775
0.15	0.1899
0.20	0.2074

Sample images of successful attacks

ϵ = 0.01



epsilon = 0.01□Misclassified as: cat

epsilon = 0.01□True Label: truck

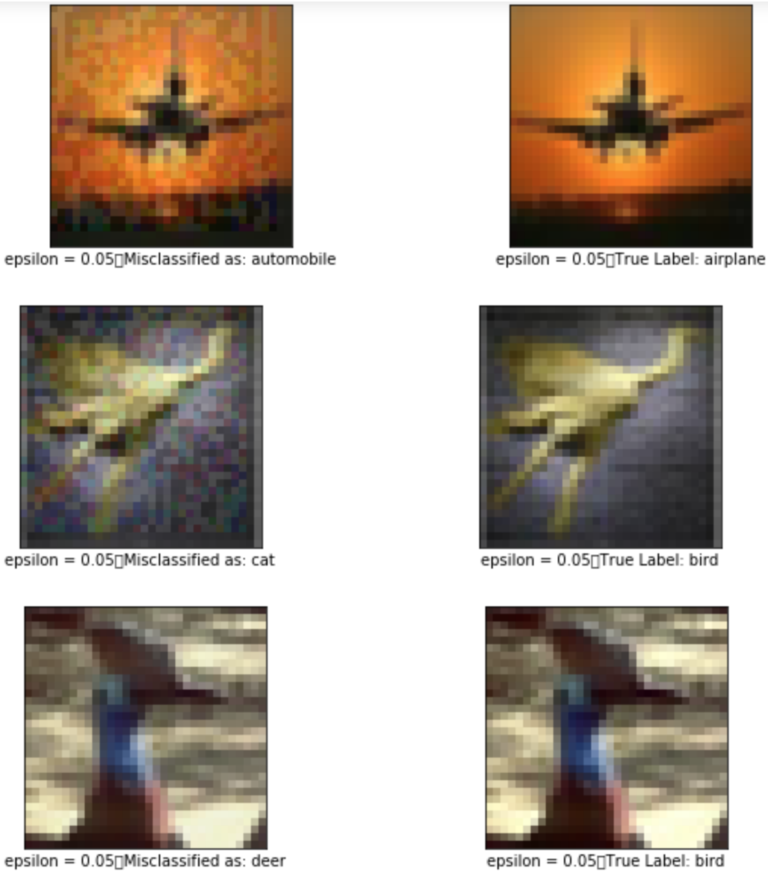
epsilon = 0.01□Misclassified as: cat

epsilon = 0.01□True Label: dog

epsilon = 0.01□Misclassified as: airplane

epsilon = 0.01□True Label: deer

ϵ = 0.05



epsilon = 0.05□Misclassified as: automobile

epsilon = 0.05□True Label: airplane

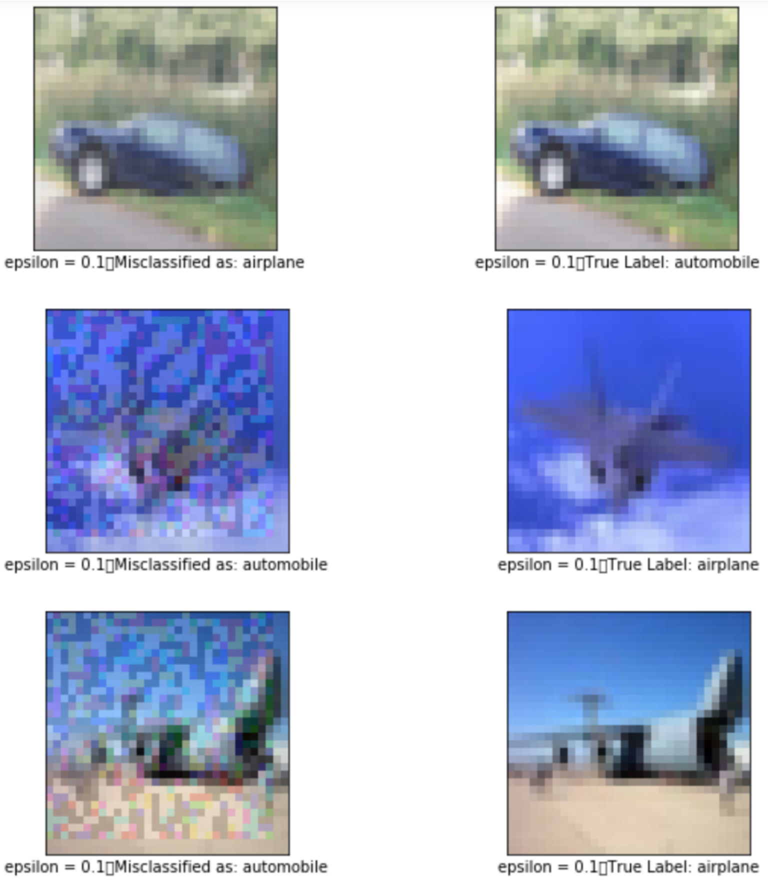
epsilon = 0.05□Misclassified as: cat

epsilon = 0.05□True Label: bird

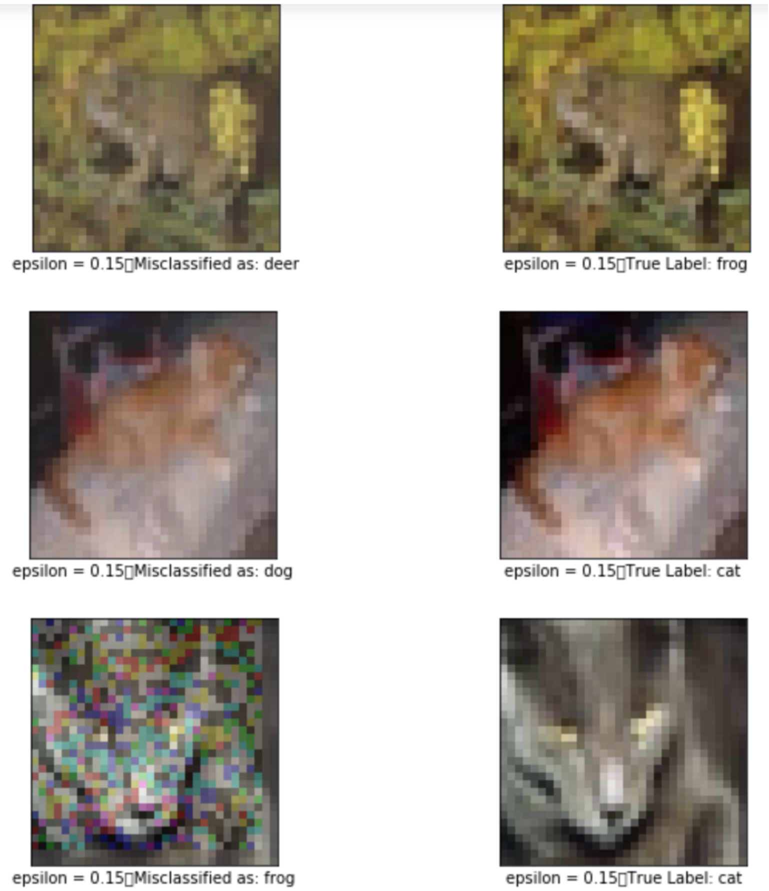
epsilon = 0.05□Misclassified as: deer

epsilon = 0.05□True Label: bird

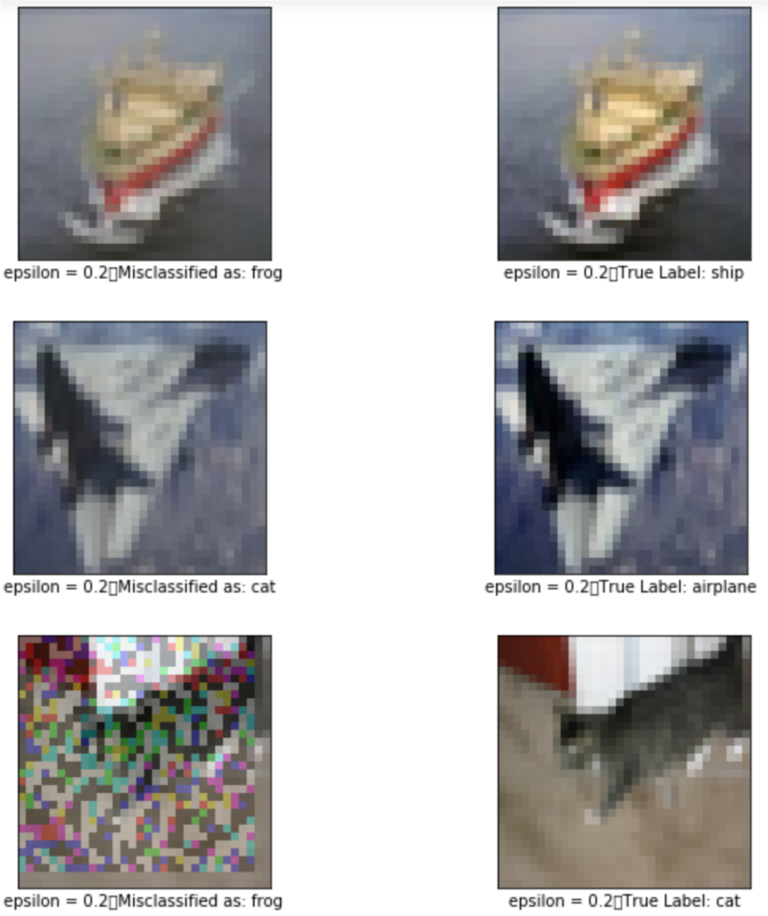
ϵ = 0.10


epsilon = 0.1□Misclassified as: airplane
epsilon = 0.1□True Label: automobile
epsilon = 0.1□Misclassified as: automobile
epsilon = 0.1□True Label: airplane
epsilon = 0.1□Misclassified as: automobile
epsilon = 0.1□True Label: airplane

ϵ = 0.15


epsilon = 0.15□Misclassified as: deer
epsilon = 0.15□True Label: frog
epsilon = 0.15□Misclassified as: dog
epsilon = 0.15□True Label: cat
epsilon = 0.15□Misclassified as: frog
epsilon = 0.15□True Label: cat

ϵ = 0.20


epsilon = 0.2□Misclassified as: frog
epsilon = 0.2□True Label: ship
epsilon = 0.2□Misclassified as: cat
epsilon = 0.2□True Label: airplane
epsilon = 0.2□Misclassified as: frog
epsilon = 0.2□True Label: cat

Iterative FGSM Method

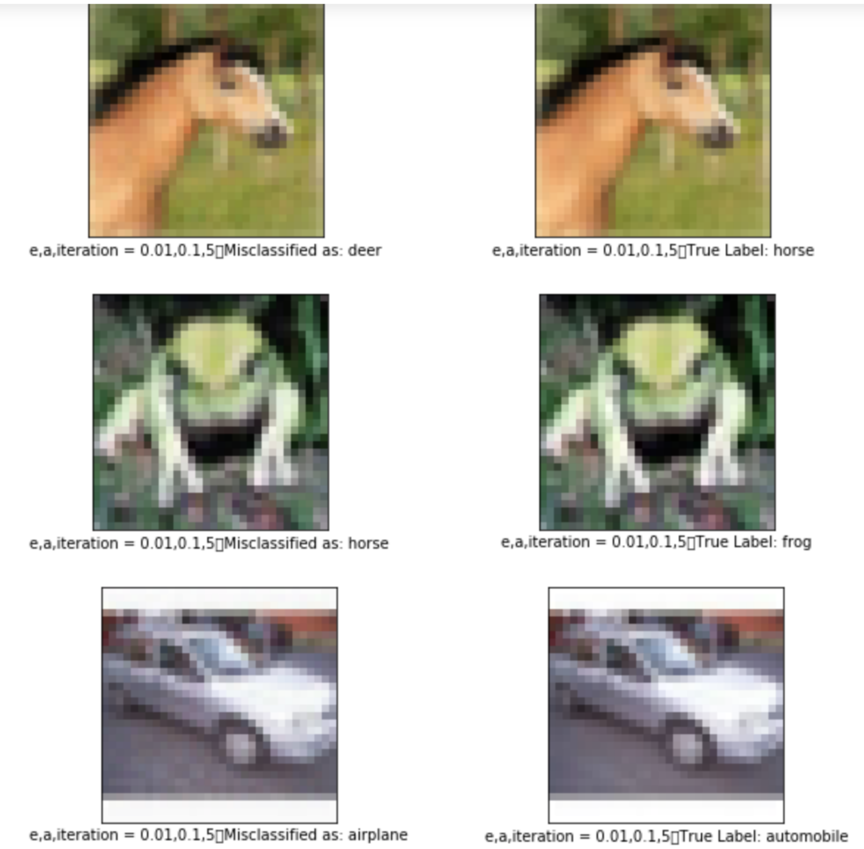
attack success for different epsilon (alpha = 0.1, iteration = 5)

Epsilon value	Alpha	Iteration	Accuracy
0.01	0.1	5	0.1482
0.01	0.1	10	0.1482
0.01	0.5	5	0.1482
0.05	0.1	5	0.1938
0.05	0.1	10	0.1937
0.05	0.5	5	0.1904
0.1	0.1	5	0.2641

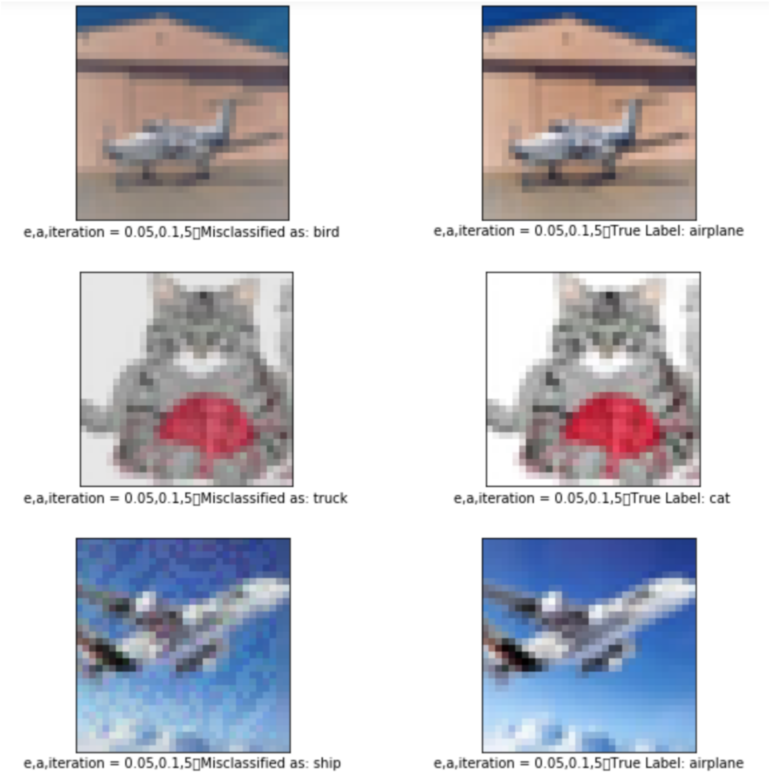
Epsilon value	Alpha	Iteration	Accuracy
0.1	0.1	10	0.2746
0.1	0.5	5	0.251
0.2	0.1	5	0.4001
0.2	0.1	10	0.4403
Untitled			

Sample images of successful attacks

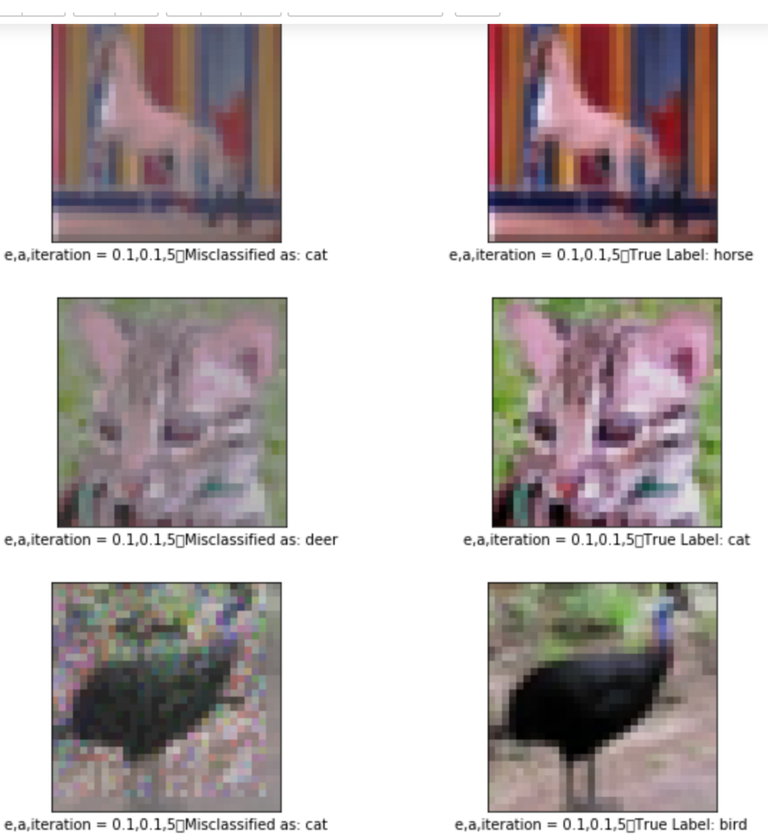
ϵ = 0.01, α = 0.1, iteration = 5


e,a,iteration = 0.01,0.1,5Misclassified as: deer
e,a,iteration = 0.01,0.1,5True Label: horse
e,a,iteration = 0.01,0.1,5Misclassified as: horse
e,a,iteration = 0.01,0.1,5True Label: frog
e,a,iteration = 0.01,0.1,5Misclassified as: airplane
e,a,iteration = 0.01,0.1,5True Label: automobile

ϵ = 0.05, α = 0.1, iteration = 5



ϵ = 0.1, α = 0.1, iteration = 5



e,a,iteration = 0.1,0.1,5□Misclassified as: cat

e,a,iteration = 0.1,0.1,5□True Label: horse

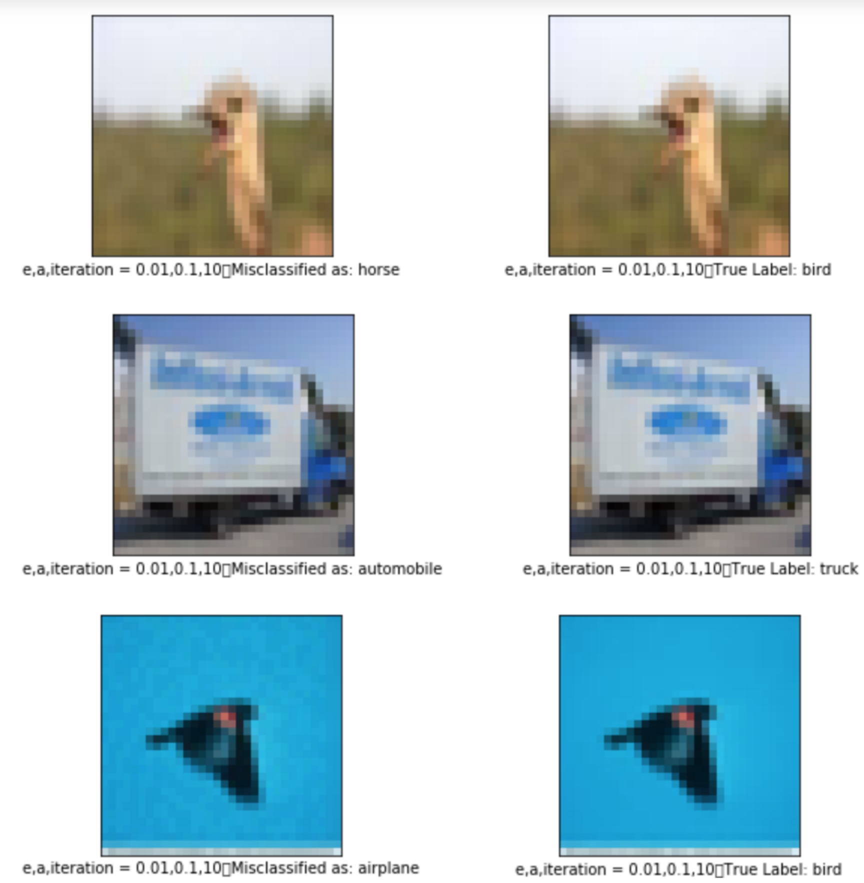
e,a,iteration = 0.1,0.1,5□Misclassified as: deer

e,a,iteration = 0.1,0.1,5□True Label: cat

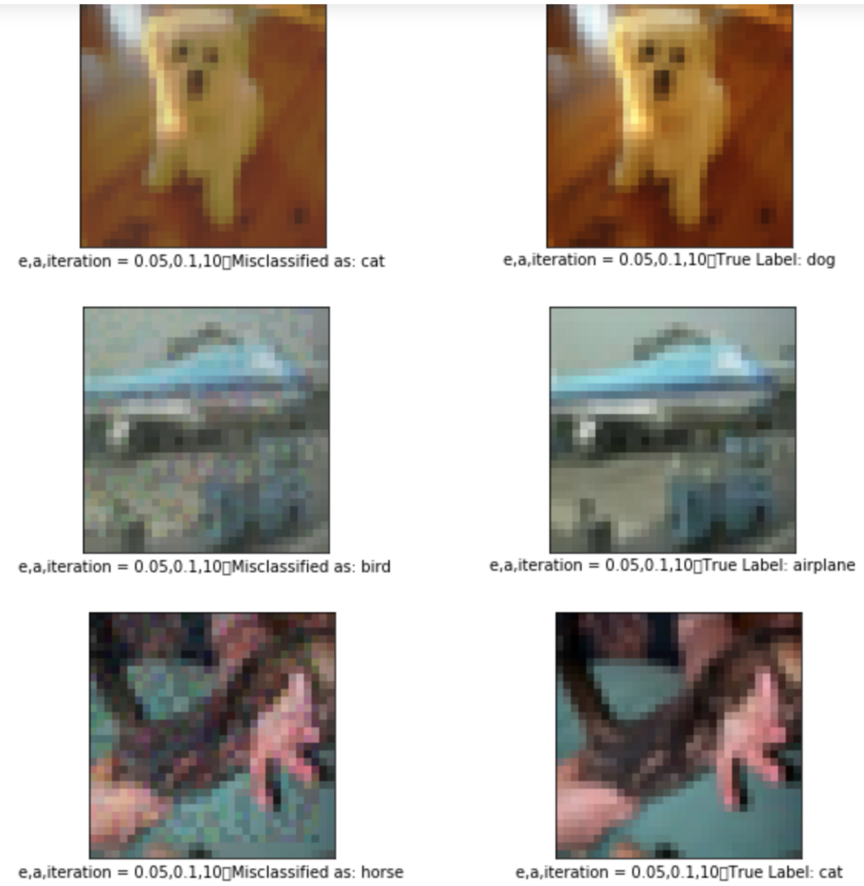
e,a,iteration = 0.1,0.1,5□Misclassified as: cat

e,a,iteration = 0.1,0.1,5□True Label: bird

ϵ = 0.01, α = 0.1, iteration = 10


e,a,iteration = 0.01,0.1,10▯Misclassified as: horse
e,a,iteration = 0.01,0.1,10▯True Label: bird
e,a,iteration = 0.01,0.1,10▯Misclassified as: automobile
e,a,iteration = 0.01,0.1,10▯True Label: truck
e,a,iteration = 0.01,0.1,10▯Misclassified as: airplane
e,a,iteration = 0.01,0.1,10▯True Label: bird

ϵ = 0.05, α = 0.1, iteration = 10



e,a,iteration = 0.05,0.1,10□Misclassified as: cat

e,a,iteration = 0.05,0.1,10□True Label: dog

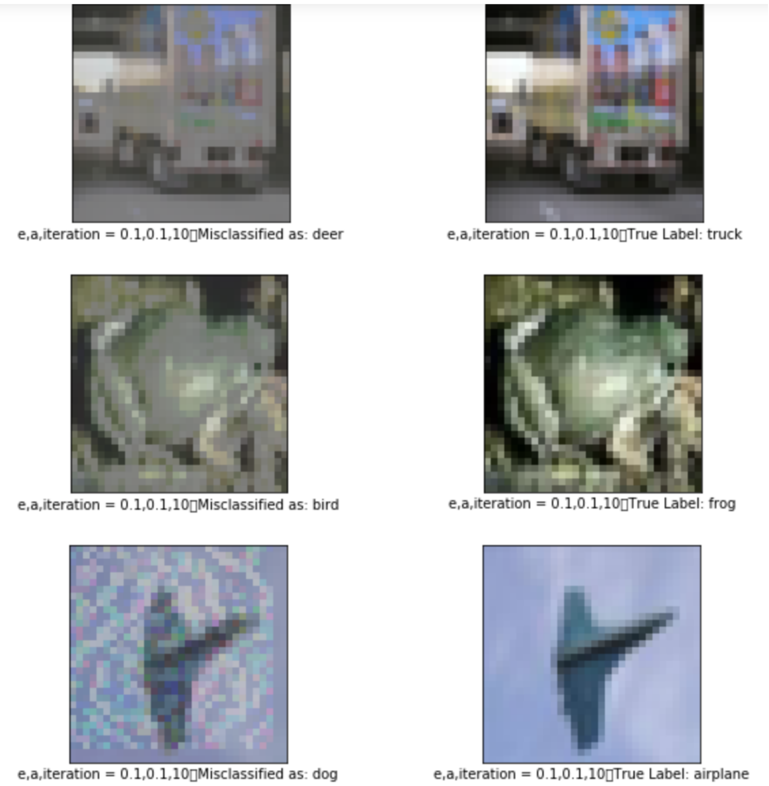
e,a,iteration = 0.05,0.1,10□Misclassified as: bird

e,a,iteration = 0.05,0.1,10□True Label: airplane

e,a,iteration = 0.05,0.1,10□Misclassified as: horse

e,a,iteration = 0.05,0.1,10□True Label: cat

ϵ = 0.1, α = 0.1, iteration = 10



e,a,iteration = 0.1,0.1,10□Misclassified as: deer

e,a,iteration = 0.1,0.1,10□True Label: truck

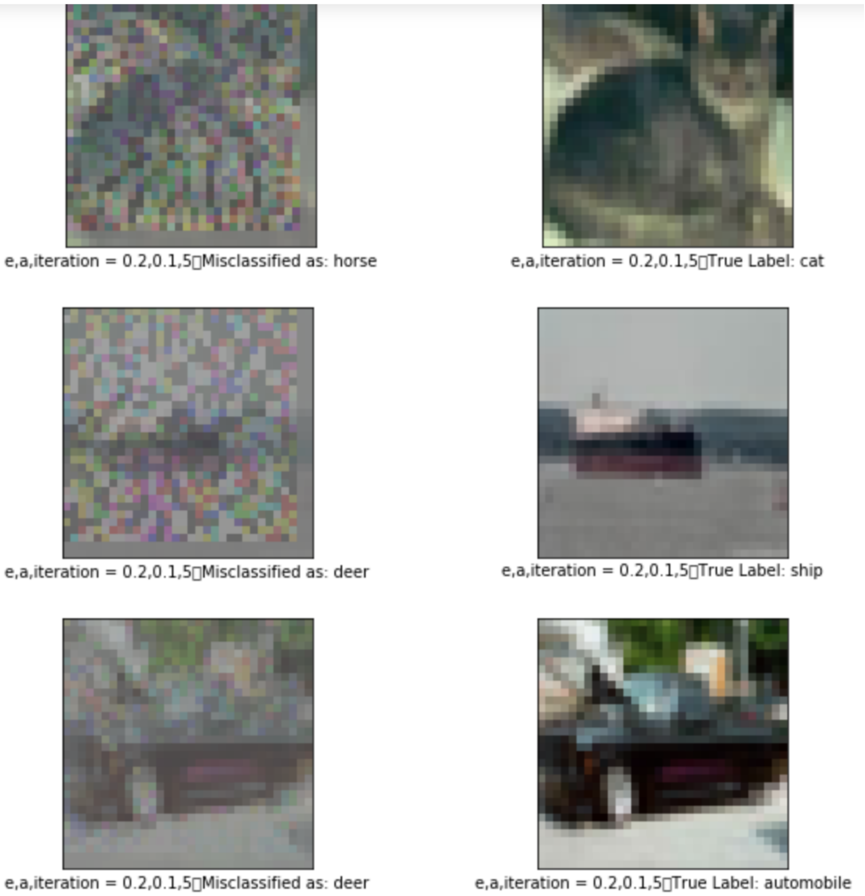
e,a,iteration = 0.1,0.1,10□Misclassified as: bird

e,a,iteration = 0.1,0.1,10□True Label: frog

e,a,iteration = 0.1,0.1,10□Misclassified as: dog

e,a,iteration = 0.1,0.1,10□True Label: airplane

ϵ = 0.2, α = 0.1, iteration = 5



e,a,iteration = 0.2,0.1,5□Misclassified as: horse

e,a,iteration = 0.2,0.1,5□True Label: cat

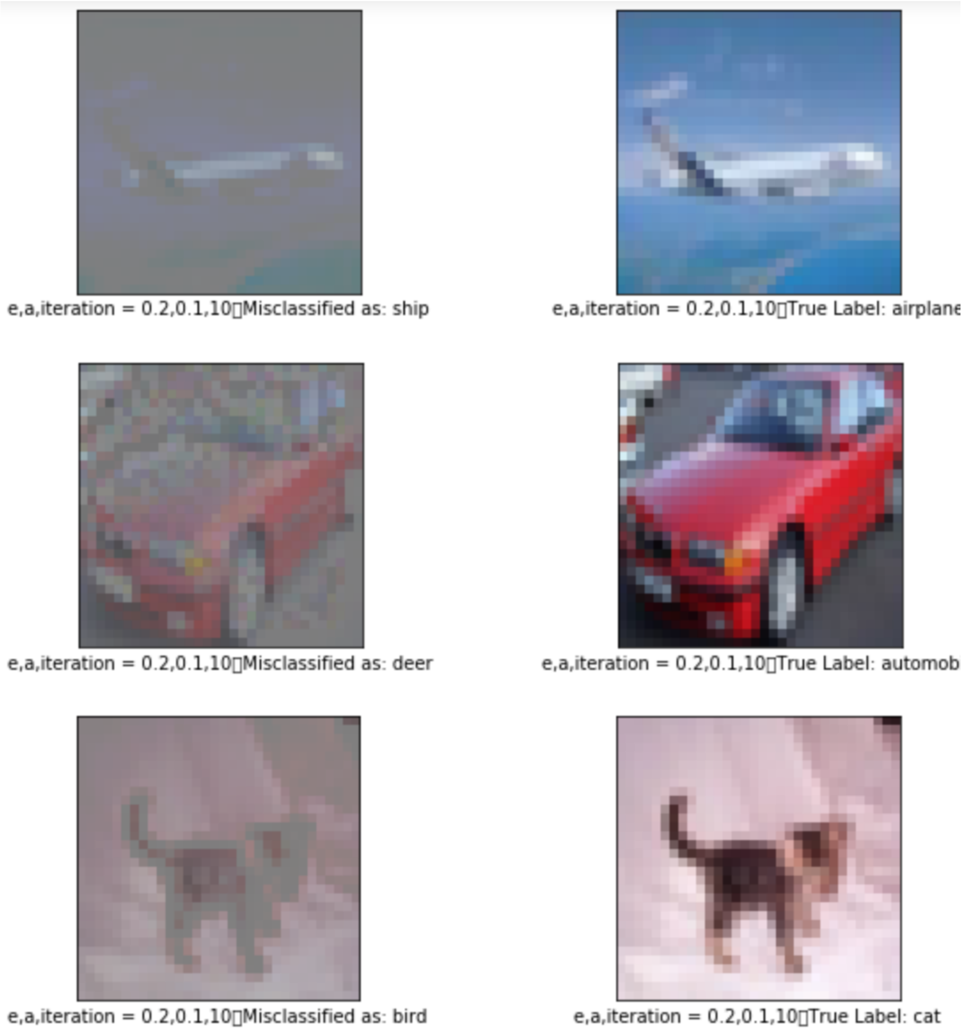
e,a,iteration = 0.2,0.1,5□Misclassified as: deer

e,a,iteration = 0.2,0.1,5□True Label: ship

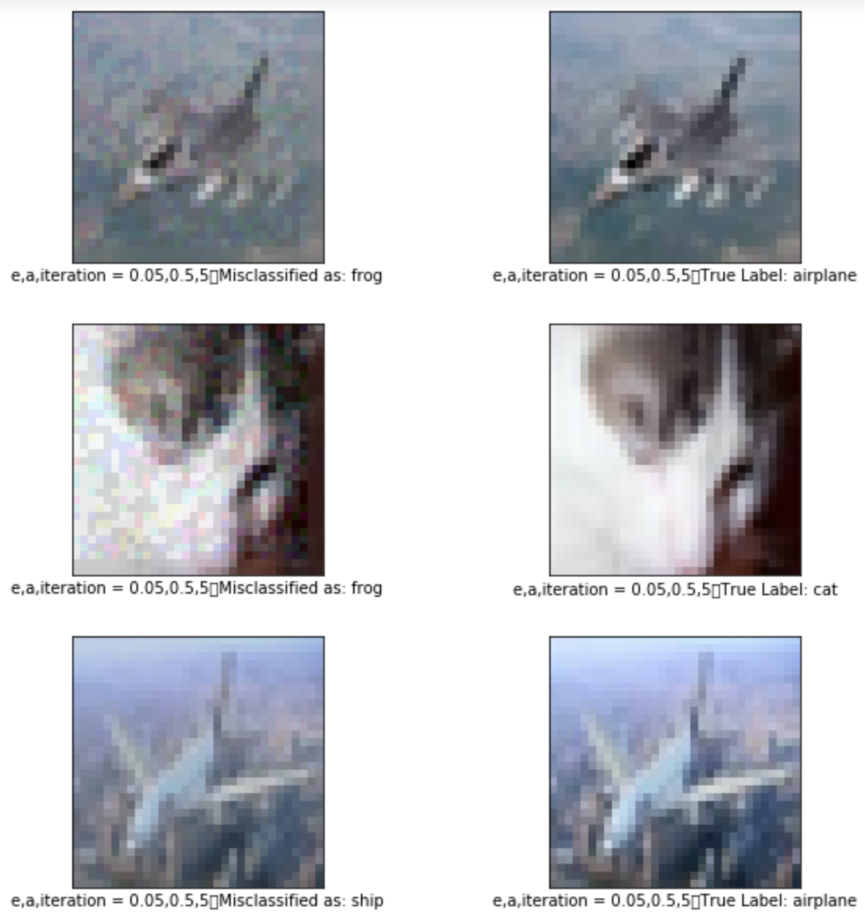
e,a,iteration = 0.2,0.1,5□Misclassified as: deer

e,a,iteration = 0.2,0.1,5□True Label: automobile

ϵ = 0.2, α = 0.1, iteration = 10


e,a,iteration = 0.2,0.1,10▯Misclassified as: ship
e,a,iteration = 0.2,0.1,10▯True Label: airplane
e,a,iteration = 0.2,0.1,10▯Misclassified as: deer
e,a,iteration = 0.2,0.1,10▯True Label: automob
e,a,iteration = 0.2,0.1,10▯Misclassified as: bird
e,a,iteration = 0.2,0.1,10▯True Label: cat

ϵ = 0.05, α = 0.5, iteration = 5



e,a,iteration = 0.05,0.5,5□Misclassified as: frog

e,a,iteration = 0.05,0.5,5□True Label: airplane

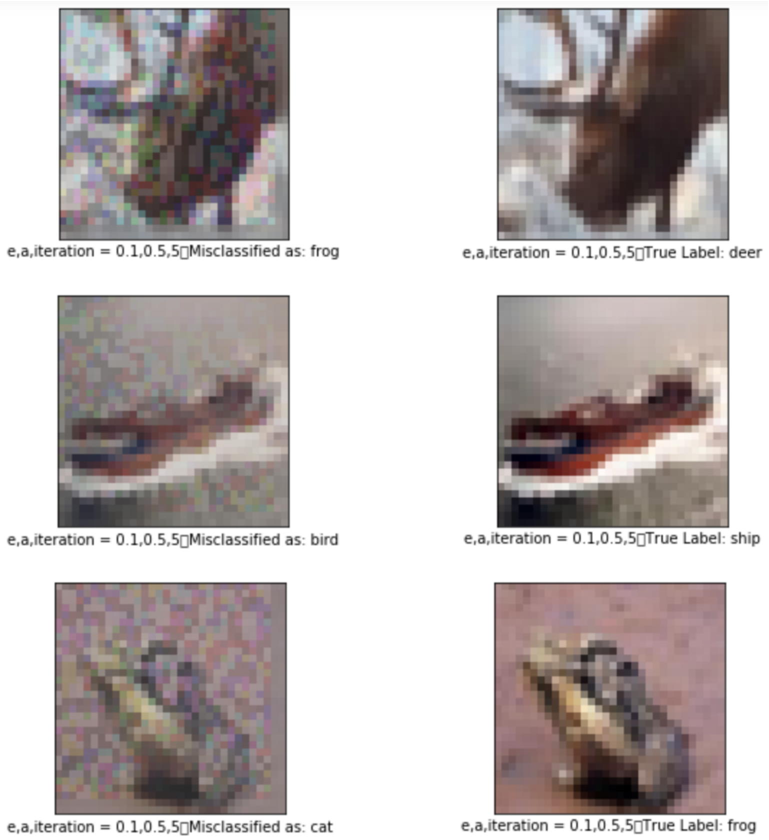
e,a,iteration = 0.05,0.5,5□Misclassified as: frog

e,a,iteration = 0.05,0.5,5□True Label: cat

e,a,iteration = 0.05,0.5,5□Misclassified as: ship

e,a,iteration = 0.05,0.5,5□True Label: airplane

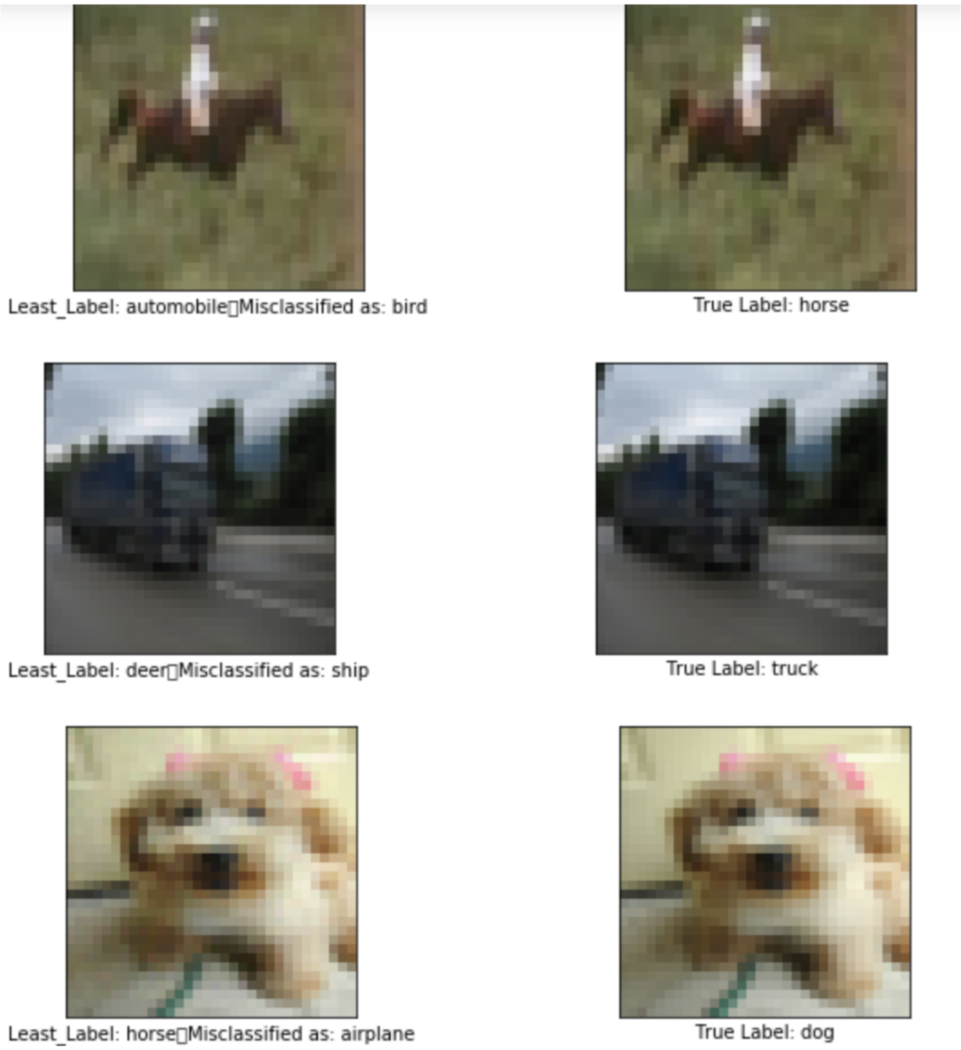
ϵ = 0.10, α = 0.5, iteration = 5


e,a,iteration = 0.1,0.5,5□Misclassified as: frog
e,a,iteration = 0.1,0.5,5□True Label: deer
e,a,iteration = 0.1,0.5,5□Misclassified as: bird
e,a,iteration = 0.1,0.5,5□True Label: ship
e,a,iteration = 0.1,0.5,5□Misclassified as: cat
e,a,iteration = 0.1,0.5,5□True Label: frog

Least-Likely FGSM

Epsilon	Alpha	Iterations	Accuracy
0.01	0.1	5	0.0228
0.05	0.1	5	0.1131
Untitled			

ϵ = 0.01, α = 0.1, iteration = 5



Least_Label: automobile□Misclassified as: bird

True Label: horse

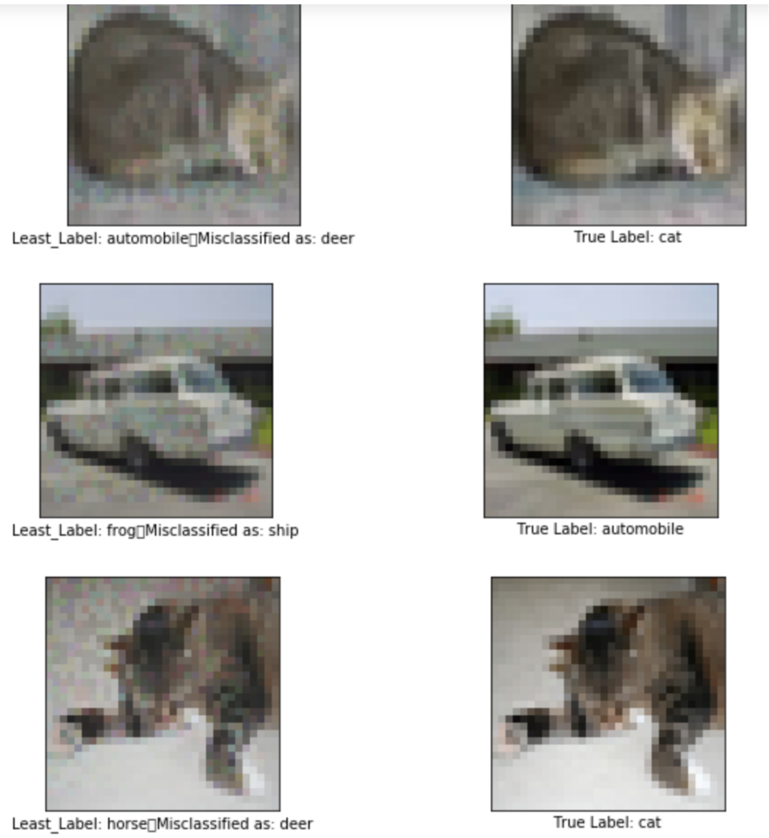
Least_Label: deer□Misclassified as: ship

True Label: truck

Least_Label: horse□Misclassified as: airplane

True Label: dog

ϵ = 0.05, α = 0.1, iteration = 5


Least_Label: automobile□Misclassified as: deer
True Label: cat
Least_Label: frog□Misclassified as: ship
True Label: automobile
Least_Label: horse□Misclassified as: deer
True Label: cat